8/22
6.8.70

THE BRITISH JOURNAL OF PSYCHOLOGY

4782

EDITED BY
JAMES DREVER

WITH THE ASSISTANCE OF
C. BURT

AIDED IN THE SELECTION OF PAPERS BY
F. C. BARTLETT L. S. HEARNSHAW W. J. H. SPROTT
E. A. BOTT M. G. KENDALL R. H. THOULESS
R. S. CREED T. H. PEAR O. L. ZANGWILL
G. C. DREW G. RYLE

ACTING EDITOR (1955): B. SEMEONOFF

VOLUME XLVI, 1955

CAMBRIDGE
AT THE UNIVERSITY PRESS
1955

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
London: Bentley House, N.W. 1
American Branch: New York
Agents for Canada, India, and Pakistan: Macmillan

[illegible]
[illegible]
Dated 6.8.70
[illegible] No. 9.22

CONTENTS OF VOL. XLVI

PART 1. FEBRUARY 1955

PAGE

Some simple methods of testing for function fluctuation. By C. C. ANDERSON . . . 1

Breadth of experience, ease of discrimination, and efficiency of generalization. By WILLIAM BEVAN and PER SAUGSTAD . . . 13

Measurement and mathematics in psychology. By A. S. C. EHRENBERG . . . 20

A theory of shape recognition. By J. A. DEUTSCH . . . 30

Motor and memory responses of neurotics and normals in the Luria association-motor technique. By A. D. B. CLARKE . . . 38

The construction of an index of intelligence. By WILLIAM A. BELSON . . . 44

Speech-breathing activity—a measure of tension and affect during interviews. By FRIEDA GOLDMAN-EISLER . . . 53

Correspondence . . . 64

Publications recently received . . . 66

PART 2. MAY 1955

Learning and retaining verbal material. By HARRY KAY . . . 81

Changes in motor response with increase and decrease in task difficulty in normal industrial and psychiatric patient subjects. By PETER H. VENABLES . . . 101

The idea of a distribution. By JOHN COHEN and C. E. M. HANSEL . . . 111

The scientific status of mental measurement. By F. W. WARBURTON . . . 122

Body-cathexis and personality. By SIDNEY M. JOURARD and PAUL F. SECORD 130

The Dvorine pseudo-isochromatic plates. By AGNES CRAWFORD . . . 139

Publications recently received . . . 144

Other publications received . . . 162

Editorial note . . . 163

Erratum . . . 164

PART 3. AUGUST 1955

On the limits of experimental psychology. By B. A. FARRELL . . . 165

The idea of independence. By JOHN COHEN and C. E. M. HANSEL . . . 178

Problem-solving as dependent on availability of functions. By PER SAUGSTAD 191

PAGE

A study of the problem of intellectual deterioration in adolescents and young adults. By K. LOVELL 199

Adaptation to level of difficulty in intelligence testing. By A. W. HEIM . 211

A test of the significance of the difference between the means of correlated measures, based on a simplification of Student's t. By JOSEPH SANDLER 225

Correspondence 227

Editorial note 230

Publications recently received 231

PART 4. NOVEMBER 1955

Obituary Notice: SIR GODFREY THOMSON, 1881–1955 245

A comparison of the personality profile (16 P.F.) of eminent researchers with that of eminent teachers and administrators, and of the general population. By R. B. CATTELL and J. E. DREVDAHL 248

Subjective judgements—some experiments employing experienced and inexperienced observers. By R. G. HOPKINSON 262

Changes in time perception induced by an anaesthetic drug. By HANNAH STEINBERG 273

Some factors influencing the effectiveness of an instructional film. By S. LANER 280

Freud and the culture-psychologists. By TERENCE WHITE GERVAIS . . 293

A study of racial attitudes during and after the war by the Thurstone technique. By A. B. FITT 306

Notes on the Shaw test. By D. B. BROMLEY 310

Publications recently received 312

Other publications received 324

INDEX OF AUTHORS

PAGE
Anderson, C. C. 1
Belson, W. A. 44
Bevan, W. 13
Bromley, D. B. 310
Cattell, R. B. 248
Clarke, A. D. B. 38
Cohen, J. 111, 178
Crawford, A. 139
Deutsch, J. A. 30
Drevdahl, J. E. 248
Ehrenberg, A. S. C. 20
Farrell, B. A. 165
Fitt, A. B. 306
Gervais, T. W. 293
PAGE
Goldman-Eisler, F. 53
Hansel, C. E. M. 111, 178
Heim, A. W. 211
Hopkinson, R. G. 262
Jourard, S. M. 130
Kay, H. 81
Laner, S. 280
Lovell, K. 199
Sandler, J. 225
Saugstad, P. 13, 191
Secord, P. F. 130
Steinberg, H. 273
Venables, P. H. 101
Warburton, F. W. 122

INDEX OF PUBLICATIONS REVIEWED

Aarek, W. *From Loneliness to Fellowship* 80
Adcock, C. J. *Factorial Analysis for Non-mathematicians* 158
Ahmavaara, Y. *Transformation Analysis of Factorial Data* 232
Alexander, F. & Ross, H. *Dynamic Psychiatry* 73
Alexander, T. *The Adult–Child Interaction Test* 320
L'Année pyschologique (1953). 68
Baynes, H. G. *The Mythology of the Soul* 315
Bellak, L. *The Thematic Apperception Test and the Children's Apperception Test in Clinical Use* 156
Bender, L. *A Dynamic Psychopathology of Childhood* 234
Bendix, R. & Lipset, S. M. *Class Status and Power* 78
Benoit, H. *The Many Faces of Love* 316
Bergmann, G. *The Metaphysics of Logical Positivism* 233
Bernard, H. W. *Psychology of Learning and Teaching* 152
Biddle, W. E. *Integration of Religion and Psychiatry* 323
Braatøy, T. *Fundamentals of Psychoanalytic Technique* 148
Brand, H. *The Study of Personality* 315
Burton, J. (ed.). *Group Discussion* 238
Buytendijk, F. J. J. *Traité de psychologie animale* 70
Callewaert, H. *Graphologie et Physiologie de l'Ecriture* 157
Carroll, J. B. *The Study of Language* 72
Cartwright, D. & Zander, A. *Group Dynamics* 75
Chambers, E. G. *Psychological Tests for Accident Proneness and Industrial Proficiency* 321
Cooper, L. F. & Erickson, M. H. *Time Distortion in Hypnosis* 239
D.S.I.R. & M.R.C. *First Report the Joint Committee on Human Relations in Industry* 79
Destunis, G. *Einführung in die Medizinische Psychologie* 318
Dunbar, F. *Emotions and Bodily Changes* 235
The Durham Research Review (No. 5) 161
Durkin, H. E. *Group Therapy for Mothers of Disturbed Children* 238
Duyckaerts, F. *La notion de normal en Psychologie Clinique* 318
Education & Psychology (Mental Testing number) 323
Ehrenstein, W. *Probleme der ganzheitspsychologischen Wahrnehmungslehre* 231
Evans, J. *Three Men* 160
Eysenck, H. J. *The Psychology of Politics* 154
Farrell, B. A. (ed.). *Experimental Psychology* 313
Fenichel, H. & Rapaport, D. (eds.). *The Collected Papers of Otto Fenichel* 148
Ferm, V. *A Dictionary of Pastoral Psychology* 242
Festinger, L. & Katz, D. *Research Methods in the Behavioural Sciences* 67
De Forest, I. *The Leaven of Love* 78
Freud, S. *The Interpretation of Dreams* 315
Friedmann, M. V. *Olive Schreiner* 242
Gengerelli, J. A. & Kirkner, F. J. (eds.). *The Psychological Variables in Human Cancer* 236
Gerth, H. & Mills, C. W. *Character and Social Structure* 154
Gregory, C. C. L. & Kohsen, A. *Physical and Psychical Research* 231
Gringes, W. W. *Laboratory Instrumentation in Psychology* 158
Grygier, T. *Oppression* 322
Guilford, J. P. *Psychometric Methods* 145
Havighurst, R. J. & Neugarten, B. L. *American Indian and White Children* 321
Heath, R. G. *Studies in Schizophrenia* 147
Holsopple, J. Q. & Miale, F. R. *Sentence Completion* 157
Hovland, C. I., Janis, I. L. & Kelley, H. H. *Communication and Persuasion* 71
Huisman, D. *L'Esthétique* 159
Jersild, A. T. *Child Psychology* 314
Jones, M. *Social Psychiatry* 317

Jung, C. G. *Answer to Job* 242

Jung, C. G. *The Development of Personality* 149

Kainz, F. *Physiologische Psychologie der Sprachvorgänge* 70

Kelly, E. L. & Fiske, D. W. *The Prediction of Performance in Clinical Psychology* 150

Kenny, J. P. & Pursuit, D. G. *Police Work with Juveniles* 160

Kentucky Symposium: *Learning Theory, Personality Theory and Clinical Research* 147

King, H. E. *Psychomotor Aspects of Mental Disease* 237

Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach Technique* (Vol. I) 72

Kutash, S. A. & Gehl, R. H. *The Graphomotor Projection Technique* 73

Lapiere, R. T. *A Theory of Social Control* 153

Lee, R. S. *Psychology and Worship* 323

Lewis, H. *Deprived Children* 150

Lorimer, F. *Culture and Human Fertility* 240

Lowenfeld, M. *The Lowenfeld Mosaic Test* 155

McCarthy, D. J. & Corrin, K. M. *Medical Treatment of Mental Disease* 236

McNemar, Q. *Psychological Statistics* 231

Maslow, A. H. *Motivation and Personality* 146

Mays, J. B. *Growing up in the City* 321

Misiak, H. & Staudt, V. *Catholics in Psychology* 159

Oppenheim, A. N. *The Function and Training of Mental Nurses* 243

Osgood, C. E. *Method and Theory in Experimental Psychology* 144

Palas, R. *Die Frustration in persönlichkeitspsychologischer Beleuchtung* 233

Pear, T. H. *English Social Differences* 321

Proceedings of the Association for Research in Nervous and Mental Disease (14 and 15 *December* 1951) 76

Quenouille, M. H. *The Design and Analysis of Experiment* 69

Reitman, F. *Insanity, Art and Culture* 78

Roback, A. A. *Destiny and Motivation in Language* 314

Robb, J. H. *Working-class Anti-Semite* 241

Robinson, M. F. & Freeman, W. *Psychosurgery and the Self* 234

Rogers, C. R. & Dymond, R. F. *Psychotherapy and Personality Change* 237

Sarason, S. B. *Psychological Problems in Mental Deficiency* 76

Schafer, R. *Psychoanalytic Interpretation in Rorschach Testing* 319

Seward, G. H. *Sex and the Social Order* 153

Singer, E. *Personality in Handwriting* 157

Sluckin, W. *Minds and Machines* 159

Snyder, F. W. & Pronko, N. A. *Vision with Spatial Inversion* 313

Spier, J. *The Hands of Children* 240

Stern, E. (ed.). *Die Tests in der Klinischen Psychologie* 158

UNESCO. *The University Teaching of Social Sciences: Sociology, Social Psychology and Anthropology* 155

Viteles, M. S. *Motivation and Morale in Industry* 79

Walker, K. & Fletcher, P. *Sex and Society* 241

Wellisch, E. *Isaac and Oedipus* 77

White, E. *Christian Life and the Unconscious* 241

Winch, R. F. *The Modern Family* 80

Wisdom, J. O. *The Unconscious Origins of Berkeley's Philosophy* 66

Wolfenstein, M. *Children's Humor* 152

Woodward, M. *Low Intelligence and Delinquency* 240

Woodworth, R. S. & Schlosberg, H. *Experimental Psychology* 68

Wright, G. G. N. *The Teaching of Arabic Numerals* 161

Wulfleck, J. W. & Bennett, E. M. *The Language of Dynamic Psychology as related to motivation research* 320

Zangwill, O. L. *Psychology as the Study of Behaviour* 312

Zubek, J. P. & Solberg, P. A. *Human Development* 233

VOLUME XLVI FEBRUARY 1955 PART 1

478270

SOME SIMPLE METHODS OF TESTING FOR FUNCTION FLUCTUATION

BY C. C. ANDERSON

Department of Education, University of Cambridge

I. *The concept of function fluctuation* (pp. 1–3). II. *Early methods of measuring function fluctuation* (pp. 3–6). III. *Thouless's correlation of differences* (pp. 6–8). IV. *Finney's analysis of variance model* (pp. 8–9). V. *Experimental study* (pp. 9–10). VI. *Conclusion* (p. 11). *References* (pp. 11–12).

I. THE CONCEPT OF FUNCTION FLUCTUATION

Thouless (1936) made three important contributions to the study of function fluctuation: first, he coined the term 'function fluctuation' itself; secondly, he was the first psychologist clearly to articulate its meaning in relation to test error; and thirdly, he devised methods by which the extent of its activity could be measured. He saw that lack of perfect self-consistency between parallel tests of the same mental function was determined partly by test error—the fact that the groups of items making up the two tests are not exactly measuring the same thing because of inevitable bias in the selection from the item population—and partly by a real variability in the function being measured, a tendency for which he proposed to reserve the term 'function fluctuation'.

> We may define the reliability of any measuring instrument as its degree of success in giving the same reading when applied to equal quantities. It may be tested by seeing how far the measuring instrument gives identical readings when applied twice to the same unvarying object....Absence of self-consistency in test results can, however, only be taken as an indication of unreliability as above defined if repetition of the test with the same individuals is really a measurement of the same unvarying mental function. This is not necessarily the case. Let us suppose that the ability, or whatever else we are attempting to measure, is in the same individual quantitatively different at different times. The fact that our test gives different results at different times is then no evidence that it is unreliable (that it gives different readings when used to measure quantities which are actually equal). Different test results would now follow even from the use of a perfectly reliable test, because the quantities measured are really different at different times of testing. (Thouless, 1936, pp. 325–6.)

The concept is not synonymous with crude variability in performance described in such statistics as Hull's (1927) variance of observed scores, or Hertzman's (1938) sum of all possible differences between scores or ranks. This would appear an elementary point, but recently Gulliksen (1950, p. 197) has confused the two because of an ambiguity in terminology. 'Methods of determining the instability of a trait as distinct from the instability of a test have been suggested by...Thouless (1936, 1939), Preston (1940), and Jackson & Ferguson (1941)'. However, the other authors cited are dealing with crude variability. Jackson & Ferguson are concerned entirely with obtaining a measure of test reliability in terms of analysis of variance instead of by the logically and statistically inferior method of correlations, and their gamma statistic is based on a comparison between the size of the errors of measurement and the size of the differences between the individuals tested. Preston attacks the situation in which subjects have taken a number of the same tests, or parallel tests of the same mental function, on different testing occasions. To obviate the labour of calculating for each individual the variance of his test scores, which might be taken as an index of what Preston calls 'trait variability', Preston (1940, p. 215) derives, in his own terms: 'an expression for the average variance of the distribution of individual scores from test to test, and an expression for the variance of these variances in terms of the number of tests and the intercorrelations between them'. Gulliksen's confusion comes in part from Preston's belief that he is measuring trait variability, a term borrowed from Paulsen (1931) who uses it to mean function fluctuation, whereas Preston is in fact measuring crude variability. The two concepts are, of course, related, but are not identical.

Another semantic point, illustrating a possible ambiguity in Thouless's concept, has been raised by Finney (1953), who states that he prefers the term 'performance fluctuation', on grounds which he does not clearly specify. Of course it is obvious that what is hypothesized abstractly as mental functions can only be measured in terms of mental products. E. L. Thorndike (1924, p. 228) indicated this many years ago: 'All scientific measures of intelligence are measures of some product produced by the person or animal in question, or the way in which some product is produced.' In this sense function fluctuation is a performance fluctuation since we cannot measure mental functions directly. However, the use of a term like product or performance fluctuation encourages its possible confusion with crude variability, a confusion which Thouless very properly avoided by his use of the convenient 'fiction' or 'function' fluctuation.

It may be that Finney wishes to claim the concreteness of pseudo-behaviourism for his phrase, but this will only be at the expense of meaning. The justification for such 'mentalistic' concepts as 'function' is that they account most reasonably for observed facts (e.g. that the within-day reliability coefficients are markedly, or at least consistently, greater than the between-day coefficients), and they are hypotheses in Ayer's (1936) sense, and 'testable explanatory-cum-predictive assertions' in Ryle's phrase (1949, p. 25). Ryle defends mental concepts as explanations of behaviour characteristics. He suggests that 'when we describe people as exercising qualities of mind, we are not referring to occult episodes of which their overt acts and utterances are effects; we are referring to those overt acts and utterances themselves' (*ibid.* p. 25). Again: 'When a person is described by one or other of the intelligence-epithets such as "shrewd" or "silly", "prudent" or "imprudent", the description imputes to him not the knowledge, or ignorance, of this or that truth, but the ability, or inability, to do certain sorts of things' (*ibid.* p. 27). And finally: 'It is being maintained throughout this book that when we characterize people by mental predicates, we are not making untestable inferences to any ghostly processes occurring in streams of consciousness which we are debarred from visiting; we are describing the ways in which those people conduct parts of their predominantly public behaviour' (*ibid.* p. 51).

Burt (1950, p. 164) defends his phrase 'the inheritance of intelligence' in much the same way:

> But most psychologists...would nowadays be quite happy to accept the view of the logical positivists, and say that, in psychology as in other sciences, induction does not seek to prove general laws of causation, but merely to provide 'propositional functions'—logical models, in fact—from which verifiable singular propositions could be readily derived. Thus in claiming that intelligence is inherited, we are not to be understood as putting forward a new generalization about the 'concrete constitution' of the mind... but merely as formulating a compact and valuable scheme for particular predictions about particular individuals or particular groups of individuals.

And finally Tolman[1] (1932, p. 144) finds it necessary to introduce mental concepts into his system. He

> conceives mental processes as functional variables intervening between stimuli, initiating physiological states, and the general heredity and past training of the organism, on the one hand, and final resulting

---

[1] The author is indebted to Dr Thouless for this reference.

> responses on the other. These intervening variables...are to be discovered, in the last analysis, by behaviour experiments. They have to be inferred 'back' from behaviour. They are precipitated out from the empirical correlations which can be observed between specific stimuli and initiating physiological states, on the one hand, and specific resultant acts, on the other. They are to behaviour as electrons, waves, or whatever it may be, are to the happenings in inorganic matter. There is nothing private or 'mentalistic' about them. They are pragmatically conceived, objective variables, the concepts of which can be altered and changed as proves most useful.

Finney also wishes to distinguish 'function fluctuation' from 'incentive fluctuation', and this yields some clue to his quarrel with Thouless's terminology. For Thouless the concept 'function fluctuation' is a description of a kind of behaviour and does not imply any psychological theory of the nature of mental functions or of their neurological correlates—it is what Cattell (1946) might term a descriptive 'surface' trait. The weather, state of health, strength of moods, incentives, boredom, each might be a source of fluctuation in different individuals on the same testing or the same individual on different testing occasions, either by serving to concentrate or deflect attention, or by surcharging or reducing the amount of mental energy available, or by any other means which need not necessarily be amenable to such metaphorical description. The term 'function fluctuation' implies neither its own determinants nor the mechanism by which these causal factors work. Finney, however, appears to imagine that the term implies that the cause of fluctuation lies within the function itself, which naturally fluctuates and would do so without, and over and above, the impact of any external or subjective stimulus. If function fluctuation were used in Finney's restricted sense, then it would be possible to distinguish between function fluctuation and incentive fluctuation, and indeed between function fluctuation and many other kinds, such as temperature fluctuations, mood fluctuation, health fluctuation, and so on. However, the term, as used more reasonably by Thouless, stands for a fact which could have any of the above causes.

II. EARLY METHODS OF MEASURING FUNCTION FLUCTUATION

The early methods of measuring function fluctuation stem clearly from a line of research which was concerned, quite mistakenly, with 'validating empirically' the Spearman-Brown 'prophecy' formula for predicting the increase in reliability to be expected from a given increase in the length of a test. It was rapidly discovered that the split-half coefficient, corrected for double length by the Spearman-Brown formula, yielded coefficients markedly, or at least consistently, higher than reliability estimates derived by correlating alternate forms of a test given on different testing occasions, and it was widely believed—Holzinger (1923*b*), Lanier (1927), Jordan (1935) and many studies occurring *passim* in the *Journal of Educational Psychology*—that this discrepancy was caused by the effect of function fluctuation, active between testing occasions, in reducing the alternate forms coefficients, while leaving unscathed the corrected split-half. While this may not be a completely valid explanation—there are spurious factors, notably that of the timing of the test (Thouless, 1939), which increase the size of the split-half and would do so even if functions did not fluctuate between tests—it did encourage psychometrists to believe that the discrepancy between perfect correlation and the alternate forms correlation could be considered as a measure of the activity of function fluctuation and test error, while the difference between unity and the split-half or corrected split-half might be taken as a measure of the strength of test error alone. Accordingly, a comparison of the two coefficients might yield some measure of the influence of function fluctuation.

This is the position taken up by Paulsen (1931, p. 218):

> Present statistical theory does not anticipate any difference in the reliability coefficients calculated from measurements made all at one time or distributed over a period of time. However empirical results have shown that test-retest and the odd-even coefficients do differ markedly in size. The discrepancy indicates that the trait or function being measured is relatively 'variable'...changes in the performance of the individuals tested are responsible for the discrepancy between reliability coefficients based on adjacent trials (odd-even) and those separated by a time-interval of a week or a month (test-retest).

Paulsen formed the idea of correcting the test-retest coefficient for the attenuation due to test error, using the split-half reliabilities in the denominator of the attenuation formula, so that any discrepancy between the revised estimate and unity would measure the extent of the influence of function fluctuation. The revised estimate itself was to be termed a coefficient of trait variability (C.T.V.).

Quite apart from the criticism offered by Thouless that it would be more logical to use the difference between unity and the revised estimate as a C.T.V., it is doubtful whether the test-retest coefficient is to be preferred to the alternate forms estimate as a measure of between-occasions reliability, and the use of the split-half estimates as measures of test error appears inadvisable.

The idea of measuring the strength of function fluctuation by comparing between-day and within-day estimates of test reliability was carried a stage further by Dunlap (1933), who proposed to use Spearman's tetrad type of factorial analysis to determine whether 'quotidian variability', a term borrowed from Woodrow (1932) and identical with Paulsen's use of trait variability and Thouless's of function fluctuation, is active in the function being measured. 'When the tetrads from the four halves of two forms, each form given at a different time, equal zero, the evidence is that individual variability is negligible and that the Spearman-Brown formula will give exactly the same value as the correlation between two forms' (Dunlap, 1933, p. 448). Thouless has noted that the tetrad method is formally equivalent to his own, but he would prefer the safer course of using the within-day alternate forms correlation instead of Dunlap's split-halves as an estimate of test reliability affected by test error only.

Pursuing Dunlap's ideas, Ferguson (1941*a*) used Holzinger's bi-factor method of analysis to factorize the reliability coefficients obtained from M.H.T. 21, 23 and 26, which were administered to a sample of 212 children of 11+ in West Yorkshire schools at intervals of 1 week. The three tests were similar in content and regarded as parallel. Each was split into two (odd and even halves) and each half was correlated with every other. The matrix of intercorrelations then consisted of three correlations between halves of tests given on the same day, and twelve correlations between halves of tests given on different days. From the analysis of the matrix one general and three group factors emerged, each group factor clustering round the two halves of a test taken on the same day. From this Dunlap would have concluded that there was evidence for the activity of function fluctuation, but Ferguson notes that there is another possible explanation of the discrepancies between the within-day and between-day correlations—spurious factors such as the presumed correlation of test error and the speeding of the test. His conclusion—that either explanation is feasible and both together probable—demonstrates that his experiment has little psychological merit in that the experimental design, with its use of split-halves, completely precludes the possibility of a decision between the two explanations which were appreciated by the author before the analysis had begun, and which had been evidently established and discussed in the research literature. Had Ferguson used parallel tests within any particular day as an estimate of test error instead of the split-half, and had the same results been obtained, then it might have been safely concluded that function fluctuation was active in the function under measurement.

The factorial method, originally suggested by Dunlap, seems not to have been used extensively in testing for function fluctuation. Yet its value to this end, especially when a large number of repetitions are available, is obvious. Consider the well-known decrease in the reliability coefficient with increase in time-interval between testing. R. L. Thorndike (1933), in his summary of the effect of the interval between test and retest on the constancy of the Stanford-Binet I.Q., demonstrated, in Ferguson's (1941 *b*, p. 125) phrase, 'that the correlation between Binet test and retest varies as some inverse function of the time-interval separating the successive testings'. Heim & Wallace (1949) administered Heim's A.H. 4 eight times to nine subjects and intercorrelated each day's scores with every other. The resulting matrix showed that the correlations calculated between scores on adjacent days are consistently higher than those calculated between scores on days farther apart. Thus the columns of correlations representing the middle testing occasions will show a smaller variance than those representing the earlier and later occasions, and the highest correlations always lie close to the diagonal cells of the matrix. Babington Smith (1950, p. 28) factorized this matrix, obtaining a large general factor and, inevitably, a small bipolar contrasting the first with the last four occasions. His conclusion runs: 'The first factor is heaviest on the middle tests, not, I suggest, because there is a factor of intelligence which is most strongly represented there, but because the situation of taking a test has undergone progressive change, and the middle tests are not so different from any other test as a test at one end is from a test at the other end. The loadings on the second, bipolar, factor are associated with the order in which the tests are given and the progressive changes.' However, if we see the matrix as contrasting within- and between-adjacent days' correlations, then Babington Smith's 'progressive change in the test situation' becomes the progressively active influence of function fluctuation or perhaps a time-trend with fluctuations, and the explanation of the term 'different' in the phrase, 'the middle tests are not so different from any other tests', is that functional variability has not had time to act so extensively in the middle tests. He suggests (*ibid.* p. 79) that 'the illustration [Heim's data] suits me because it is one of the few cases I know where we have material which will bring out the effects of serial links on the factor-structure'. Rather the data appears to be not a criticism of factor-analysis but an illustration of one of its possible uses. The fact that the matrix contains evidence of function fluctuation which factor-analysis records is a ground for faith in the value of the factorial method.

A caveat should be entered here against the validity of any interpretation of the second and third factors, which are probably insignificant since so few people took the tests. Babington Smith does not appear to have tested the significance of each residual matrix, and does not explain why he stopped factorizing when he did. Should the last two factors prove to be insignificant, much of the point of each interpretation is lost. However, this would not affect the present writer's case since the decrease in the reliability coefficient with increased time-interval between testing is a well-established psychometric fact, and tests with larger numbers of subjects would doubtless show the same pattern of intercorrelations as that presented by Heim.

Of course it may be agreed that the test-retest correlation is not a reliability coefficient in that there is the possibility of direct memory transference which encourages a correlation of test error. Kelley (1924, p. 202) says of the test-retest: 'We may call such a situation one in which there is a correlation between errors, meaning that whatever elements of uncertainty or chance operated in the solution of the first question, this would tend to operate in the same manner in the solution of the second.' However, against this can be placed the fact that any correlation of test error affects all the coefficients and may not, therefore, be used as an explanation of differences between them. Functional variability or rather functional trend is the most reasonable explanation of the pattern of Heim and Wallace's correlations, although the design of their experiment is such that no definite proof is afforded. In any case it seems clear that factorial analysis could be a useful method of testing for evidence of function fluctuation if parallel forms of a test were administered on each testing occasion so that a reasonable estimate of the extent of test error could be made.

Woodrow (1932), in trying to test for the existence of what he called 'quotidian variability', substituted for the correlation, with its logic of similarity, the standard deviation with its logic of differences, and suggested, without apparently being aware of the then-novel method of analysis of variance, that a comparison of the extent of variability between testing occasions to the average variability within testing occasions might yield a good index of 'quotidian variability' for a subject repeating the test. Thouless has demonstrated how Woodrow's clumsier statistic can be replaced by the more elegant analysis of variance.

III. THOULESS'S CORRELATION OF DIFFERENCES

Thouless derived his correlational method of testing for evidence of function fluctuation and measuring its amount from early British psychometric research, specifically Spearman's (1904*a*, *b*, 1910) attempt to deal with the problem of errors of measurement in psychological data, errors which attenuate the hypothetical 'true' correlation between true error-free scores on the correlated variables. Spearman (1904*a*, p. 88) divides the errors into two groups: 'accidental' or 'compensating' and 'systematic', 'constant', or 'non-compensating'; and proposes to correct for the attenuation, induced by the former, by a formula which necessitates the administration of two independent tests of the same mental function together on at least two different testing occasions, 'on the idea that the size of these accidental errors can be measured by the size of the discrepancies between different measurements of the same things' (Spearman, 1910, p. 274).

Starting from this background, Thouless applied this double-test-retest method to a more useful purpose. He assumes that two alternate and parallel forms of a test, A and B, are administered on two different testing occasions. Each individual has, therefore, four scores, A_1 and B_1 on the first testing occasion, and A_2 and B_2 on the next. Thus six correlations may be calculated: the within-day $r_{A_1B_1}$ and $r_{A_2B_2}$, between-day test-retest $r_{A_1A_2}$ and $r_{B_1B_2}$, and between-day alternate forms $r_{A_1B_2}$ and $r_{B_1A_2}$. Thouless states that, if test error is the only source of variation, and if errors of measurement are not correlated, then there will be no significant correlation of the difference between the test and retest scores for each of the two tests, i.e. $r_{(A_1-A_2)(B_1-B_2)}$ will not differ significantly from zero.

If, however, function fluctuation is active, then the scores of the subjects in the retesting will deviate positively or negatively, the extent and the direction of the deviations being common to A_2 and B_2. This implies that the within-day reliabilities, $r_{A_1B_2}$ and $r_{A_1B_2}$, have a common factor of 'temporal contiguity' (Thouless's term) which tends to increase their separate correlations, while relatively the alternate forms $r_{A_1B_2}$ and $r_{B_1A_2}$ show a decreased relationship. The amount $r_{(A_1-A_2)(B_1-B_2)}$, which may be termed the 'criterion' of whether function fluctuation is present, suffers from the fact that it is not independent of the size of r_{AB}, and so is not suitable for indicating the amount of fluctuation present. Accordingly, Thouless derives an index of that amount by dividing the criterion by the average of the within-day reliabilities

$$\frac{r_{(A_1-A_2)(B_1-B_2)}}{\frac{1}{2}(r_{A_1B_1}+r_{A_2B_2})}.$$

This is termed the 'double test-retest index of function fluctuation' (Thouless, 1936, p. 333), and will be designated I_1.

The method was earlier used by Brown (1910) in his attempt to disprove the assumption of no error-correlation which Yule had demonstrated to be the basis of Spearman's original and most famous correction for attenuation. Brown & Thomson (1940, p. 158) argue, however, that any significant correlation of differences is best taken as evidence not of error correlation but of 'functional variability', and Holzinger (1923*a*), using the same method, concurred with this view.

The present writer has been able to trace few researches making use of Thouless's proposals. Vernon (1937) stressed their importance. Ferguson (1941*b*) gives an account of the method but is mistaken in his assertion that a test of significance, presumably the normal standard error to the correlation, is applicable to the index. Jackson & Ferguson (1941) suggest that there is no evidence for function fluctuation with cognitive abilities without specifying any research. Connor (1952) uses the Spearman tetrad type of analysis. Remmers & Whisler (1938) note Thouless's formulation, and attempt to improve on it. They take tests U and V as measures of the same function, each being further subdivided into items, $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$, and these items so arranged that there is no constant time-difference beteeen the items of the different series—$U_1V_1U_2V_2$ or $U_1U_2V_1V_2$. They assert that the '"true" relation between U and V is given by the correction for attenuation' (Remmers & Whisler, 1938, p. 90). If all the V items are now formed into one test and all the U items into another, then r_{UV} corrected for attenuation may be computed, and the discrepancy between the two corrected correlations or between their squares may be taken as a criterion of whether function fluctuation is present. Quite apart from the fact that this method offers no advantages over Thouless's, the validity of the formula for correction for attenuation in this context is doubtful (Thouless 1939).

One point about Thouless's account appears to have gone without discussion. Thouless suggests that the index will have unity as its maximum value. Speaking of the index, he writes (Thouless, 1939, p. 333): 'Its value is zero if there is no day-to-day variation of the mental function measured by the two tests. Function fluctuation produces a positive value of the index which would be unity if function fluctuation were the sole cause of absence of complete correlation between repetition of the tests (i.e. if there were no test unreliability).' However, there is no apparent mathematical reason why the maximal value of the index should be unity, although in practice it will probably be less than unity with tests of high reliability, and the present writer has found that tests showing $r_{AB} < 0{\cdot}3$ sometimes yield indices greater than unity.

IV. FINNEY'S ANALYSIS OF VARIANCE MODEL

The use of analysis of variance in testing for function fluctuation in the repeated performances of an individual was first suggested by Thouless. Finney (1953) has recently suggested a model for use with groups of subjects. Making the usual assumptions, and provided each test on each occasion is independently standardized to the same mean and standard deviation, with q subjects being given the same set of p tests on each of r occasions, for all functions the analysis will be as shown in Table 1. The test for evidence of function fluctuation is the variance ratio $F = B/C$, but, like Thouless's criterion with which it is directly correlated, it varies with the size of r_{AB}. Finney cites three indices, two of which have value for the present research. His second index (I_2) is the ratio of the fluctuation variance to the error variance, estimated by $(B/C-1)/p$. This will have little relation to Thouless's (I_1) because it is perfectly correlated with the variance-ratio, but it has the merit of serving as a check on I_1. It may be shown[1] that Thouless's criterion is approximately equal to

Table 1. *Analysis of variance model*

Variations	D.F.	Observed mean square
Subjects (S)	$q-1$	S
Tests (T)	0	—
Occasions (O)	0	—
ST	$(p-1)(q-1)$	A
SO	$(q-1)(r-1)$	B
TO	0	—
STO	$(p-1)(q-1)(r-1)$	C
Total	$pr(q-1)$	

$$\frac{\dfrac{B-C}{p}}{\dfrac{B-C}{p}+C}\quad\frac{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)}{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)+1}=\frac{I_2}{I_2+1}.$$

These equivalences are designated approximate—Thouless's criterion is not exactly equal to $I_2/(I_2+1)$, mainly because of the heterogeneity of the standard deviations used in the calculation of the correlations.

Finney's third index (I_3) is the ratio of the fluctuation variance to the variability of subjects, estimated by

$$\frac{r\,(B-C)}{S-A-B+C}.$$

This appears to be psychologically akin to I_1. However, since it is a ratio with many terms, the size of its standard error will be considerable. Moreover, since it contains terms other than B and C, there is no statistical necessity that it should be significant even when its related criterion is significant, whereas it is psychologically fundamental that a significant criterion should give rise to a significant index. Both points may be illustrated from the approximate standard error to I_3 given to the present writer by Dr Daniels.

---

[1] These approximate equivalences were suggested to the present writer by Miss V. R. Cane of the Medical Research Council Applied Psychology Unit, Cambridge, and by Dr H. E. Daniels, of the Statistical Laboratory, the University of Cambridge.

$$\text{S.E.}_{I_2}=\frac{y\,(y+r)}{r}\sqrt{\left[\frac{2}{k}\left\{\frac{S^2+A^2}{(S-A)^2}+\frac{(B^2+C^2)}{(B-C)^2}\right\}\right]},$$

where $y=I_3$, k = number of subjects. Table 2 shows relevant data for three selected functions concerning the size of B/C, the estimate of I_3 and the standard error of I_3. Thus each of the variance ratios B/C is significant at the 5 % level (1·48) while not one of the comparable indices is significant. Accordingly, it seems wise to restrict the use of the index to its raw form and to compare indices by ranking only.

Table 2. *The significance of the criterion and related indices*

Functions	B/C	I_3	S.E.$_{I_3}$
26	2·52	1·916	1·5
25	1·594	1·087	1·67
21	2·445	11·79	43·57

V. EXPERIMENTAL STUDY

Parallel tests of twenty-six functions were administered on the same day to a group of seventy-four children in the scholarship class in two schools in Cambridgeshire, and re-administered approximately 1 week later. Each of the four distributions of scores for any function was normalized to the same mean and standard deviation, and Table 3 gives the comparative statistics for Thouless's and Finney's models.

It is clear that the two models give much the same results. However, it is obvious that I_3 has the disadvantage of being occasionally negative. This may happen when the fluctuation variance is smaller than the error variance, especially in the case of tests showing high reliability and discriminating well between objects. Such a negative index would quite validly represent—at least from the psychological point of view—a low estimate of the amount of fluctuation present. However, a negative index may also derive from the fact that in the denominator $A+B>S+C$. This condition—responsible for the negative I_3 of functions 7 and 22—is a product of the joint effects of poor test discrimination between subjects (relatively low S variance), marked differential reaction by the subjects to the different tests (relatively large ST variance) and a greater fluctuation than error variance ($B>C$). Now in Table 3, the negative I_3 indices for functions 7 and 22 are negatively correlated with their corresponding I_1 indices, which are relatively large and positive. While there is no reason to expect the rank-order coefficient between I_1 and I_3 to be unity always (since the one contains terms that the other does not), yet it should be significantly positive and high.

The solution to the problem may be that I_3 is only a good index under certain conditions. Consider it in the form

$$I_3=\frac{\dfrac{B}{C}-1}{\dfrac{S}{C}-\dfrac{A}{C}-\dfrac{B}{C}+1}.$$

If sets of random numbers rather than test scores had been chosen, the mean squares would arise simply by chance, and the ratios S/C, A/C and B/C would approximate to unity. In this case I_3 tends to approach the indeterminate form O/O which is highly unstable, and the smallest change in many of the mean squares could swing the index from minus infinity to plus infinity. The conclusion is, therefore, that I_3 is only of service if B/C is significant and A/C insignificant, assuming that S/C is also significant and that the tests do actually discriminate between pupils. The index I_3 may be a good measure of the extent of function fluctuation when it is certain that only one function is being measured, and the function fluctuation is significant. This hypothesis is borne out by a glance at the functions in Table 3, where B/C is not significant or of borderline significance (7, 10, 12, 16, 20, 22, and 23), and by considering the correlations between the test administrations, in which the tests (non-cognitive) are obviously not parallel.

Table 3. *Comparative statistics from Thouless's and Finney's models*

Function	Thouless's criterion	I_1	$F=B/C$	I_2	$\frac{I_2}{I_2+1}$	I_3
1	0·471[1]	0·512	2·752	0·876	0·462	0·0475
2	0·442[1]	0·526	2·617	0·8085	0·442	0·089
3	0·400[1]	0·435	2·384	0·692	0·408	0·024
4	0·143	0·159	1·291	0·1455	0·127	0·011
5	0·410[1]	0·500	2·374	0·687	0·407	0·078
6	0·526	0·715	3·252	1·126	0·530	0·314
7	0·154	1·247	1·337	1·685	0·144	−5·488
8	0·275[1]	0·658	1·737	0·3685	0·269	0·492
9	0·498[1]	0·868	2·98	0·99	0·497	1·331
10	0·206	0·379	1·549	0·2745	0·215	0·192
11	0·549[1]	0·692	3·571	1·2855	0·562	0·366
12	0·178	0·335	1·47	0·235	0·190	0·224
13	0·283[1]	0·590	1·862	0·431	0·301	0·486
14	0·374[1]	0·900	2·247	0·6235	0·384	0·937
15	0·371[1]	0·672	2·144	0·572	0·364	0·662
16	0·195	0·424	1·454	0·227	0·185	0·454
17	0·392[1]	0·806	2·271	0·6355	0·388	1·426
18	0·246[1]	2·149	1·667	0·3335	0·250	1·362
19	0·347[1]	1·209	1·974	0·487	0·328	2·528
20	0·090	0·548	1·227	0·1135	0·102	1·394
21	0·411[1]	1·137	2·445	0·7225	0·419	11·79
22	0·162	1·296	1·364	0·182	0·154	−1·775
23	0·210	0·520	1·503	0·2515	0·201	0·532
24	0·018	0·120	1·068	0·034	0·033	0·193
25	0·242[1]	0·785	1·594	0·297	0·229	1·087
26	0·409[1]	0·863	2·52	0·760	0·432	1·916

[1] Significance at the 5% level.

Table 4. *Correlations for functions* 7 *and* 22

Function	A_1B_1	A_2B_2	A_1A_2	B_1B_2	A_1B_2	B_1A_2
7	0·208	0·039	0·291	0·205	0·014	0·002
22	0·152	0·098	0·437	0·202	0·116	−0·083

The instability of I_3 can also be illustrated from the S.E. provided by Dr Daniels, in which, whenever $S=A$ or $B=C$, the S.E. moves quickly to infinity. It has been suggested to the present writer[1] that the form of the S.E. condemns I_3 as a suitable index.

In general, if the tests are not really parallel and if the activity of function fluctuation is not significant by the B/C test, I_3 loses precision, becomes unstable and can acquire any value, positive or negative, without meaning anything.

---

[1] By J. S. Macpherson, who gave valuable advice about this passage.

VI. CONCLUSION

The best simple methods of testing for function fluctuation, those of Thouless and Finney, give similar results, and it is suggested that they should be used in future research into the amount of fluctuation present in cognitive and non-cognitive functions. The present writer is carrying out such a programme with scholarship and pre-scholarship children, and research with different tests of many functions should be extended to other populations.

I am greatly indebted to Dr R. H. Thouless for his critical supervision of this paper, especially his pruning of the vaguer expressions. Ben S. Morris, Director of the National Foundation for Educational Research in England and Wales, suggested the topic of research, gifted some of the testing material, allowed me to consult Finney's memorandum, and has shown great interest in the progress of the work. John S. Macpherson, Research Officer of the Scottish Council for Research in Education, gave valuable linguistic and statistical advice. Finally, I must thank the Carnegie Trust for a grant in aid of the research.

REFERENCES

Ayer, A. J. (1936). *Language, Truth and Logic*. London: Gollancz.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Brit. J. Psychol.* III, 296–322.

Brown, W. & Thomson, G. H. (1940). *The Essentials of Mental Measurement*. Cambridge University Press.

Burt, C. (1950). The trend of national intelligence. *Brit. J. Sociol.* I, 154–68.

Cattell, R. B. (1946). *The Description and Measurement of Personality*. London: Harrap.

Connor, D. V. (1952). The effect of temperamental traits upon the group intelligence test performance of children. Unpublished Ph.D. thesis, University of London.

Dunlap, J. W. (1933). Comparable tests and reliability. *J. Educ. Psychol.* XXIV, 442–53.

Ferguson, G. A. (1941*a*). A bi-factor analysis of reliability coefficients. *Brit. J. Psychol.* XXXI, 172–82.

Ferguson, G. A. (1941*b*). *The Reliability of Mental Tests*. London: U.L.P.

Finney, D. J. (1953). *A Note on the Measurement of Performance Fluctuation*. Memorandum to the National Foundation for Educational Research in England and Wales.

Guilford, J. P. (1950). *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill.

Gulliksen, H. (1950). *The Theory of Mental Tests*. New York: Wiley.

Heim, A. W. & Wallace, J. G. (1949). The effects of repeatedly testing the same group on the same intelligence test. Part I. Normal subjects. *Quart. J. Exp. Psychol.* I, 151–9.

Hertzman, M. (1938). Two equations for the study of variability. *Amer. J. Psychol.* LI, 571–4.

Holzinger, K. J. (1923*a*). An analysis of errors in mental measurement. *J. Educ. Psychol.* XIV, 278–88.

Holzinger, K. J. (1923*b*). Note on the use of Spearman's prophecy formula for reliability. *J. Educ. Psychol.* XIV, 302–5.

Hull, C. L. (1927). Variability in amount of different traits possessed by the individual. *J. Educ. Psychol.* XVIII, 97–106.

Jackson, R. W. B. & Ferguson, G. A. (1941). *Studies on the Reliability of Tests*. Toronto: Bulletin 12. Department of Educational Research, the University of Toronto.

Jordan, R. C. (1935). An empirical study of the reliability coefficient. *J. Educ. Psychol.* XXVI, 416–26.

Kelley, T. (1924). *Statistical Method*. New York: Macmillan.

Lanier, L. H. (1927). Prediction of the reliability of mental tests and tests of special abilities. *J. Exp. Psychol.* X, 69–113.

Paulsen, G. B. (1931). A coefficient of trait variability. *Psychol. Bull.* XXVIII, 218–19.

Preston, M. G. (1940). Concerning the determination of trait variability. *Psychometrika*, V, 275–81.

Remmers, H. H. & Whisler, L. (1938). Test reliability as a function of method of computation. *J. Educ. Psychol.* XXIX, 81–92.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Smith, Babington (1950). Factor analysis. Part II. An evaluation of factor analysis from the point of view of a psychologist. *J. R. Statist. Soc.* B, XII, 73–85.

SPEARMAN, C. S. (1904*a*). The proof and measurement of association between two things. *Amer. J. Psychol.* XV, 72–101.

SPEARMAN, C. S. (1904*b*). 'General Intelligence' objectively determined and measured. *Amer. J. Psychol.* XV, 201–92.

SPEARMAN, C. S. (1910). Correlation calculated with faulty data. *Brit. J. Psychol.* III, 271–95.

THORNDIKE, E. L. (1924). The measurement of intelligence. *Psychol. Bull.* XXXI, 219–52.

THORNDIKE, R. L. (1933). The effect of the interval between test and retest on the constancy of I.Q. *J. Educ. Psychol.* XXIV, 543–9.

THOULESS, R. H. (1936). Test unreliability and function fluctuation. *Brit. J. Psychol.* XXVI, 325–43.

THOULESS, R. H. (1939). The effect of errors of measurement on correlation coefficients. *Brit. J. Psychol.* XXIX, 383–403.

TOLMAN, E. C. (1932). *Purposive Behaviour in Animals and Men.* New York: The Century Co.

VERNON, P. E. (1937). The Stanford-Binet Test as a psychometric instrument. *Character & Pers.* VI, 99–113.

WOODROW, H. (1932). Quotidian variability. *Psychol. Rev.* XXXIX, 245–56.

(*Manuscript received* 12 *April* 1954)

BREADTH OF EXPERIENCE, EASE OF DISCRIMINATION, AND EFFICIENCY OF GENERALIZATION

By WILLIAM BEVAN and PER SAUGSTAD

Emory University, Georgia, and the University of Oslo[1]

I. *Introduction* (p. 13). II. *Experimental design and specification of hypotheses* (pp. 13–15). III. *Experimental situation* (pp. 15–16): (i) *Materials* (p. 15); (ii) *Experimental groups* (p. 15); (iii) *Experimental sessions* (pp. 15–16); (iv) *Measures of performance efficiency* (p. 16). IV. *Results* (pp. 16–17). V. *Discussion* (pp. 17–19): (i) *Evaluation of Hypothesis* (pp. 17–18); (ii) *Subsidiary comment* (pp. 18–19). VI. *Summary* (p. 19). *References* (p. 19).

I. INTRODUCTION

Most psychological theories to the present appear to possess in common at least one important untested assumption: An organism's discriminatory mechanisms are regarded to exist as given, essentially functionally intact and ready for use, as part of its biological equipment. This has had obvious consequences. The question of representativeness of design does not present itself; the significance, for general theory, of truly developmental studies and studies of individual differences, is usually underestimated; and the adequacy of control of the subject's pre-experimental past experience—specification as a rule is limited to statements about age and *naïveté* with regard to the experiment's purpose—is generally not seriously considered.

Recently reported data (e.g. Brunswik's (1939) on the effect of varying the relative frequency of required responses; Helson's (1948) and Johnson's (1944) on the effect of varying the relative frequency of various magnitudes of stimulation; Riesen's (1950) on arrested vision in chimpanzees, Ames' (1946) on visual illusions; and Harlow's (1949) on interproblem transfer) meanwhile support the view that mature discriminative behaviour requires not only establishing some sort of functional juncture between the action of the discriminatory mechanism and that involved in the overt expression of behaviour but also the sufficient development of the discriminatory mechanism itself (cf. e.g. Hebb's (1949) developmental theory of perception, Cantril's (1950) concept of assumptive world and Harlow's (1949) of learning set). Yet little has been done, with the exception of Helson's (1948) studies of intra-organismic reference standards, to specify empirically the parameters of such a developing mechanism. Work such as Harlow's, however, suggests that there are dimensions other than those traditionally recognized that may be fruitfully explored. In a discussion of problem-solving Harlow (1951, p. 469) has commented: 'Breadth, rather than intensity of training, appears to be the key...whether we deal with subhuman or with human subjects.'

Accordingly, the aim of the present investigation has been to attempt to create a situation in which breadth of experience (training) may be conveniently given objective definition and its relation to performance efficiency in a 'new' situation evaluated.

II. EXPERIMENTAL DESIGN AND SPECIFICATION OF HYPOTHESES

The plan of the experiment consists of a combination of two traditional designs: that of the method of constant stimuli for measuring judgemental acuity and that of the transfer of training experiment. The subjects' task involved making differential judgements with one of several ranges of stimuli about a common standard on four consecutive days and then on the fifth with a non-overlapping test range. Breadth of experience is considered defined by the magnitude of the training range, novelty of situation by the difference in position of training and test ranges on the stimulus continuum. Performance efficiency is expressed in terms of the accuracy and precision of the mean judgements made with the test range.

---

[1] The present work, done at the University of Oslo during the 1952–3 academic year, represents the joint effort of both writers while the former was in residence as a Fulbright research scholar.

Since no concern was directed toward the criterion of representativeness of design, the experimenters have no wish to imply that the unidimensional definition they have applied to breadth of experience is adequate for the 'normal' condition for which one may ultimately wish to predict. Nor do they claim that the relatively invariate situation they arranged for its operation have yielded data which in any crucial respect are typical of 'normal conditions'. Representative complexity was, possibly, sacrificed to gain ease of objective specification for the key concepts.

Thus the present experimental task may lack the novelty that the laboratory discrimination problem is presumed to have for the traditionally assumed-to-be naïve white rat, dog, or rhesus monkey, for all of our subjects had already developed, probably to a highly refined degree, a concept of stimulation differential. We have rather *arbitrarily* selected the lifted-weight task as our discrimination problem, mainly because we suspect that it provides a situation requiring formal judgement met less frequently by our subjects in their pre-experiment past. We assumed, furthermore, that our subjects' skill at judging had not been developed to a degree such that it could not be further improved by the training procedures of our experiment. We also assumed, of course, that prior to this training subjects randomly grouped were on the average equally sophisticated. Harlow's conclusion quoted above was consequently transformed as follows: If subjects be allowed experience (training trials) in judging the relative heaviness of lifted weights, equal in frequency but varying in range, those with wider practice will, when confronted by weights outside the training range, be more accurate and/or precise in their discriminations. One or both types of subject should be superior in performance to those receiving no experimental training prior to presentation of the test range.

If this first hypothesis were confirmed, a second was generated from the Neo-Pavlovian theory of generalization gradients (Hull, 1943) to account for the superiority of the broader training. If subjects be allowed training experience, equal in amount and range, but differing in its similarity to the test experience (adjacency of training and test ranges on the stimulus continuum), those receiving the more similar experience will display greater accuracy and/or precision in their performance of the test task.

Should the second hypothesis lack confirmation, another means of explaining the superiority of the broader training would have to be sought. The stimulus series used for narrow and broad training differ not only in range about a common central magnitude but in the size of the inter-weight interval. This suggests that degree of distinctness of the individual training experiences is important for generalization and is reminiscent of the Köhler & von Restorff report (1933) on the relationship of 'crowding' to ease of recall, except that, of course, the Köhler & von Restorff 'crowded' and 'isolated' materials were qualitatively different. That inter-weight interval may be important is also suggested by the human-engineering studies (Chapanis, Garner & Morgan, 1949) on interval size and efficiency of dial reading. A final hypothesis was therefore formulated: If subjects be allowed training experience equal in total amount and derived from the same range within the stimulus continuum, the magnitude of the standard-variable interval in the training series will prove to be a significant factor in the accuracy and/or preciseness of their test performance. In the present instance the superiority of the wider interval training group is predicted.

III. EXPERIMENTAL SITUATION

(i) *Materials*

The weights employed as stimulus material were opaque plastic cylinders 5 cm. in diameter and 8 cm. high filled with a combination of lead shot, saw-dust and cotton. They were made up in four training series and a test series. Training series A, the long-range, wide-interval series, consisted of five weights, each weight differing from the next heaviest in the series by 40 g. and with the middlemost serving as the standard: 220 g., 260 g., 300 g. (S), 340 g., 380 g. Training series B, the short-range, narrow-interval series, like series A consisted of five weights with the standard at 300 g. The interval between adjacent weights of 10 g., however, effected a range and interval size one-fourth that of series A: 280 g., 290 g., 300 g. (S), 310 g., 320 g. Training series C, also a short-range, narrow-interval series, is the adjacent series. Both its lower end and its midpoint are closer to those of the test series than are the corresponding magnitudes in series A and B. Its interval size is proportionately equal to that of series B: 205 g., 212·5 g., 220 g. (S), 227·5 g., 235 g. Training series D is the long-range, narrow-interval series. Constructed about a midpoint of 300 g., its range is equal to that of series A, its standard-variable differential never exceeding the maximum differential of series B. This was accomplished by arranging sixteen weights into five overlapping short-range, narrow-interval sub-series: 220 g., 230 g., 240 g. (S), 250 g., 260 g.; 250 g., 260 g., 270 g. (S), 280 g., 290 g.; 280 g., 290 g., 300 g. (S), 310 g., 320 g.; 310 g., 320 g., 330 g. (S), 340 g., 350 g.; 340 g., 350 g., 360 g. (S), 370 g., 380 g. Finally the test series consisted of nine weights 4 g. apart about a standard of 100 g.: 84 g., 88 g., 92 g., 96 g., 100 g. (S), 104 g., 108 g., 112 g., 116 g.

A three-sided, roughly 'Z' shaped wooden screen, with side panels 40 cm. by 50 cm. and middle panel 50 cm. square, and with a semicircular aperture through which the subject's lifting hand could be inserted at the bottom centre of the middle panel, was devised in order to eliminate visual cues during testing. The weights were presented on a rubber-backed felt pad to minimize the possibility of auditory cues.

(ii) *Experimental groups*

Ninety-six persons, seventy-nine medical and seventeen natural science students at the University of Oslo, all in their early twenties, served as subjects. None had had experience in or knowledge of weight-lifting experiments prior to his participation. Each subject was randomly assigned to one of five groups. Those assigned to Group 1 attended only one experimental session during which they were presented with the test series. Members of the other four groups received training in weight discrimination with one of the four training series prior to test. The unequal numbers in the groups resulted from a few subjects withdrawing their participation before completion of the training sequence.

(iii) *Experimental sessions*

The training sessions, which required about 15 min. each, were held on four consecutive days. They were followed on the fifth day by the test session, also about 15 min. in duration. During each session each subject made eighty comparative judgements (Group 5, which received training series D made eighty-two on each training day), half with the standard presented first, half with the order reversed, in random sequence. The random alternation of the temporal order in which the standard was presented on successive trials was, of course, executed with the hope of distributing the bias contributed by the psychophysical time error. During the training sessions the standard (standards in the case of series D) was presented with itself as frequently as with each other weight. In the test series it was never presented with itself. Subjects, all trained and tested individually, were requested to lift with the non-dominant hand in an effort to reduce to some extent the effect of weight-experience gained before the experiment. *The testing procedure, that followed* in the method of constant stimuli, is described in the following instructions to the subject:

> This is an experiment in perception which aims at demonstrating how the ability to form judgements of a certain type develops. For this purpose we shall ask you to judge the weight of various objects.
>
> Which is your dominant hand? Put the *non-dominant* hand through the aperture of the screen before you. The wrist ought to be well inside the aperture. Now I will present to you a number of objects which are alike in every respect except their weight. You will be given one object at a time. When you feel the object against the tip of your fingers grasp it as you would a glass.
>
> With regard to the manner in which you are to lift the object, there are certain matters to which you must pay careful attention. Lift every object only once, raising it quickly and putting it down immediately. Then give your judgement. When you lift, the whole arm from the wrist up should rest on the table; in other words, you are to lift only from the wrist. Be careful that you raise the whole hand and do not rest any of your fingers on the table while you are lifting.
>
> You will be given pairs of objects one after the other. When you have lifted both, you are to tell if you think one is heavier or lighter than the first. If the last one is lighter, you say, 'light'; if it is heavier, you say, 'heavy'.

(iv) *Measures of performance efficiency*

The relative frequencies of 'heavier than' judgements for each standard-variable comparison of the test series were first combined by means of Spearman's summation method (Woodworth, 1938) to yield a mean heaviness score and standard deviation for each subject. The cut-tail correction was applied when appropriate. The mean heaviness score is comparable (assuming a normal distribution of responses) to the P.S.E. gained graphically and provides information on the degree of correspondence between the objective midpoint of the stimulus range and its subjective equivalent. It is taken as an *index of accuracy*. The standard deviation provides information about the consistency of a subject's performance about his own mean. It is taken as an *index of preciseness*. After these indices had been derived for each subject, they were used as raw scores in group comparisons. The indices entered in columns 3 and 4 of Table 1 are thus arithmetic averages of the individual accuracy and precision indices. A third proficiency score, the *index of uniformity*, was also used. This is the standard deviation of the *mean* accuracy index and indicates the extent to which subjects within each experimental group are alike in accuracy.

IV. Results

Table 1 presents the mean indices of accuracy and preciseness as well as the uniformity scores for the several experimental groups; Table 2 presents the extent to which the group means deviated from perfect accuracy; Table 3 presents the results of the various possible intergroup comparisons of preciseness and uniformity. Examination of Table 1, column 3, and Table 2 indicates almost perfect accuracy in all groups except Group 4, that trained on the adjacent series, which shows a small but significant tendency toward underestimation. The performance of Group 2, trained on the long-range, wide-interval series is reliably more precise than that of any of the other training groups. These last-mentioned, meanwhile, do not differ reliably from those receiving no formal practice before the test. Finally, there are clear differences among the groups in uniformity. As in the case of precision, Group 2 is the superior group. Groups 1, 4 and 5 do not differ significantly among themselves. Group 3, the short-range, narrow-interval group, is inferior to all other groups.

Table 1. *Mean indices of accuracy and preciseness, index of uniformity expressed in grams for the several experimental groups*

Experimental group	No. subj.	Mean index of accuracy	Mean index of preciseness	Index of uniformity
(1) No training	21	100·15	11·66	2·51
(2) Training series A	19	99·94	9·18	0·97
(3) Training series B	19	100·31	11·96	4·64
(4) Training series C	20	101·14	10·92	2·29
(5) Training series D	17	100·60	12·17	2·45

Table 2. *Summary of 't' tests to determine if experimental groups differed reliably from perfect accuracy*

Difference between mean accuracy index and perfect accuracy for each group		*t*	*P*
(1)	0·15	0·27	>0·05
(2)	0·06	0·26	>0·05
(3)	0·31	0·26	>0·05
(4)	1·14	2·19	<0·05
(5)	0·60	0·98	>0·05

Table 3. *Summary of intergroup comparison with regard to preciseness and uniformity*

Groups compared	D.F.	Diff.	*t*	*P*	Diff.	*t*	*P*
1–2	39	2·48	3·39	<0·01	1·54	3·67	<0·01
1–3	39	0·30	0·39	>0·05	2·13	2·45	<0·05
1–4	40	0·74	0·87	>0·05	0·22	0·41	>0·05
1–5	37	0·51	0·63	>0·05	0·06	0·11	>0·05
2–3	37	2·78	3·66	<0·01	3·67	4·58	<0·01
2–4	38	1·74	2·07	<0·05	1·32	3·22	<0·01
2–5	35	2·99	3·74	<0·01	1·48	3·22	<0·01
3–4	38	1·04	1·19	>0·05	2·35	2·76	<0·01
3–5	35	0·21	0·25	>0·05	2·19	2·49	<0·05
4–5	36	1·25	1·38	>0·05	0·16	0·28	>0·05

V. DISCUSSION

(i) *Evaluation of hypothesis*

With these data in mind, we may review and evaluate the status of the hypotheses specified above. The greater preciseness of Group 2 as compared to Group 3 and their lack of difference in accuracy add up to partial confirmation of the first hypothesis. This confirmation is strengthened by the greater intra-group uniformity of Group 2. Comparison of the performance of these groups with that of Group 1 indicates that the superiority of Group 2 represents a facilitating effect of this type of training. The failure of Group 3's performance to surpass that of the no-training group (Group 1) in precision, and the high-level accuracy of all groups suggests the extent to which *pre-experimental* heaviness experience benefits judging in the present experimental situation. If practice is required to match the performance of Group 3, it could have been provided for the no-training group only during the test session itself. The superiority of Group 2 over

Group 3, assuming it not to reflect sampling error, meanwhile indicates that even with relatively sophisticated subjects, types of training may be employed which will effect greater proficiency in certain aspects of performance. The relative lack of uniformity in Group 3 meanwhile suggests that some kinds of training may interfere with rather than facilitate the operation of the discriminatory mechanism.

Comparison of Group 3 with Group 4 reveals no significant difference in preciseness but the inferiority of Group 4 in accuracy. Meanwhile Group 2 retains its superiority in both preciseness and uniformity. This means a rejection of our second hypothesis and points to the inadequacy of the similarity theory as it is usually defined for discrimination experiments of this sort. The excellence of the long-range training group (Group 2) cannot be attributed to the fact that the training series contains weights which are closer in the stimulus continuum to the test series than are corresponding items of the short-range series.

Finally, comparison of Groups 2 and 5 provides a test of our third hypothesis. The significantly smaller average preciseness score and uniformity score mean confirmation with regard to these criteria. At the same time they mean a rejection of Harlow's 'breadth hypothesis' as we have specified it for this experiment. Group 2 is superior, not because its pre-test weight lifting experience covered a wider range of weights, but because it included paired items more widely separated on the weight continuum and thus more readily discriminable. Breadth of experience or breadth of training, should this experiment be representative of the requirements of at least the simplest discrimination problems, is most adequately defined, not in terms of total range of magnitudes presented to the subject, but in terms of the range of differences in magnitudes provided in the sequence of comparisons. Furthermore, the fact of Group 2's greater preciseness and uniformity as compared with the narrow-interval training groups on the narrow-interval test task prompts the suggestion that ease of discrimination appears not to be dictated solely by the size of the interstimulus interval presented to the subject but to be determined also by the extent to which his discriminatory mechanisms have been developed at the time of test. Indeed, it is not so much a matter of establishing the superiority of wide versus narrow training intervals as it is of determining the optimally discriminable interval relative to a given level of development. The present results on interval size parallel those of Lawrence (1952) reported while this experiment was in progress. The fact that his results were obtained with white rats rather than human subjects and a visual-intensity discrimination rather than a weight-discrimination task suggests that our preciseness data are independent of the present particular conditions of experiment. It also suggests the possible fruitfulness of further exploration of the relation between ease of discrimination (interval size) and transfer efficiency and its explanation both as a point of reorientation in learning theory and as a basis of more effective teaching technique.

(ii) *Subsidiary comment*

The present accuracy data have implications for Helson's (1948) and Johnson's (1944) studies of the position at which the subject 'zeros in' on the stimulus scale. Their results have shown that this statistic does not always stand in constant relation to the objective standard, but may, instead, vary with changes in the amount and magnitude of previous stimulation and with present properties of the context or ground. Our accuracy scores, meanwhile, do not conform to what might be expected from a consideration of their findings. If the reference standard (adaptation level) represents an average of all instances of stimulation, then the mean accuracy scores for Groups 2, 3 and 5 should be equal but of greater magnitude than those of Groups 1 and 4. The failure of this prediction to be confirmed may be explained in several ways without invalidating adaptation level theory. Perhaps the most likely explanation is to be found in the temporal distribution of weight experience in the training groups. Helson's and Johnson's reports apparently are based on judgements made during a single experimental session. Our training and test sessions were separated by approximately 24 hr. The results of Group 4 may be accounted for if we assume that more similar (adjacent) stimuli exert an influence upon the weighted average of experience smaller in magnitude than that of the less similar (assuming both categories to occur with equal frequency) but one of longer duration. If the Helson and Johnson shifts are only temporary effects, then the degree of permanence with which they may be induced must be further studied, as must the trend toward high accuracy in the subjects of this experiment.

VI. Summary

The relation of breadth of past experience to effectiveness of generalization was studied in the context of the classical weight-lifting experiment. Groups were trained about a standard of 300 g. with a wide-range, wide-interweight-interval series (Group 2), a narrow-range, narrow-interval series (Group 3), and a wide-range, narrow-interval series (Group 5). Another group was given a narrow-range, narrow-interval series about a 220 g. standard (Group 4) and still another (Group 1) made judgements only with the intermediate-range, narrow-interval *test* series about a standard of 100 g. Groups 2, 3 and 5 did not differ reliably from perfect accuracy in test performance. Group 4, however, displayed a small but significant positive constant error. Group 2 was, meanwhile, significantly more precise than the other groups, which did not differ from each other in preciseness. The greatest intra-group uniformity was found in Group 2, the least in Group 3. Superiority in test performance was thus found to be associated with interval size rather than breadth of training or greater adjacency of training and test weights on the stimulus continuum.

References

Ames, A. (1946). *Some demonstrations concerned with the origin and nature of sensations.* A laboratory manual (preliminary draft).

Brunswik, E. (1939). Probability as a determiner of rat behaviour. *J. Exp. Psychol.* xxv, 175–97.

Cantril, H. (1950). *The Why of Man's Experience.* New York: Macmillan.

Chapanis, A., Garner, W. R. & Morgan, C. T. (1949). *Applied Experimental Psychology*, ch. 5. New York: Wiley.

Harlow, H. F. (1949). The formation of learning sets. *Psychol. Rev.* lvi, 51–65.

Harlow, H. F. (1951). Thinking. In H. Helson (Ed.), *Theoretical Foundations of Psychology*, ch. 10. New York: van Nostrand.

Hebb, D. O. (1949). *Organization of Behaviour.* New York: Wiley.

Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.* lv, 297–313.

Hull, C. L. (1943). *Principles of Behaviour*, ch. 12. New York: Appleton Century.

Johnson, D. M. (1944). Generalization of a scale by the averaging of practice effects. *J. Exp. Psychol.* xxxiv, 425–36.

Köhler, W. & von Restorff, H. (1933). Analyse von Vorgängen in Spurenfeld. *Psychol. Forsch.* xviii, 299–342.

Lawrence, D. H. (1952). The transfer of a discrimination along a continuum. *J. Comp. Physiol. Psychol.* xlv, 511–16.

Riesen, A. H. (1950). Arrested vision. *Sci. Amer.* clxxxiii, 16–19.

Woodworth, R. S. (1938). *Experimental Psychology*. ch. 17. New York: Holt.

(*Manuscript received 23 February* 1954)

MEASUREMENT AND MATHEMATICS IN PSYCHOLOGY[1]

By A. S. C. EHRENBERG

Institute of Psychiatry (*Maudsley Hospital*), *London, S.E.*5

I. *Introduction* (pp. 20–21). II. *The spurious distinction between quality and quantity* (pp. 21–23). III. *The irrelevance of additive units* (pp. 23–25). IV. *The use of mathematics* (pp. 25–26). V. *The standardization and comparability of measurements* (pp. 26–28). VI. *Summary* (p. 28). *References* (p. 29).

1. INTRODUCTION

Psychologists tend to feel uncertain and apologetic about the numerical methods which they use in describing their observations. Mental test scores, ratings, sensory assessments and the like are often said not to be 'measurements' at all, and the validity of their mathematical analysis is doubted. Criticisms tend to be rather vague—some logical or even metaphysical attributes of 'measurement' are apparently thought to be lacking. More specific charges refer to the 'arbitrary' nature of the scoring-units, and to their alleged lack of 'equality' and 'additivity'. The consequences are implied to be dire, even if they are left obscure.

Criticisms arise because the measurements of physics, still called the 'science of measurement' even nowadays, are taken to be the necessary model of *all* measurement. But our familiarity with the conventional 'dimensions' and 'units' of physics makes us clothe them with a reality going far beyond the demonstrable empirical facts. Historically it would be inept, and factually wrong, to regard the degree scales of temperature, for instance, as derived from our present-day knowledge of 'absolute temperatures', 'equal amounts of work done per degree rise of temperature', and so on. Temperature scales were set up by the arbitrary, or better perhaps, the deliberate, subdivision of the change in something as 'far-fetched' as the length of a column of mercury; subsequent formulations have only led to equally conventional re-definitions of the scale in terms of changes in platinum resistances, thermo-couple e.m.f's, black-body radiations, and so on.

The importance of any 'theory of measurement' considerations may perhaps best be judged from physics. Here the physicist himself is generally said to have no knowledge of the 'theory of measurement', nor does he seem to have much use for it. Indeed, there is no accepted theory of measurement. Typically, the British Association Committee (1938, 1940) 'On the possibility of quantifying sensory events' could not agree on the meaning of measurement, and this is still the common state of affairs (cf. Bartlett, 1950). The various attempts which are made to develop some such theory appear doomed, like other contributions to the philosophy of science, to be ignored by the practising scientist. Recent papers on measurement in the psychological literature (e.g. Gulliksen, 1946; Coombs, 1950, 1951, Coombs, Raiffa & Thrall, 1954; Burke, 1953) lean on the better-known writings of the physicist Campbell (1928) and the psychologist Stevens (1951) which will be discussed below. It was also to these two authors that reference was made in discussions of measurement theory at the recent symposium on Flavour Assessment (1953). There the usual attitude to non-physical measurement was made explicit by Dr J. O. Irwin (1953) who, on the one hand, said that the derivation and use of certain scoring scales seemed to him to be 'thoroughly scientific' and, on the other hand, stressed the doubts that might be felt in treating such scales as if they had an 'underlying metric'. (Compare also Bartlett's comment on the earlier symposium (1950)—'When Theory whispers low "you can't", Practice replies "I must".') At the flavour symposium, Dr J. M. Shewan and the present writer (Ehrenberg & Shewan, 1953) tried to bring the matter to a head by suggesting that 'if any attention is to be paid to those people who say that one cannot "measure" in the present context for some reason or other, they should be asked to state specifically what they mean, in terms that are capable of being tested empirically'. There seems to be no instance of an actual mathematical analysis of numerical assessments which has been criticized purely on grounds of 'theory of measurement' criteria.

---

[1] This article is based on a paper read at the British Psychological Society's Annual Conference, Nottingham, April 1953, and arises partly out of work carried out on behalf of the Department of Scientific and Industrial Research (Food Investigation Organization).

Nothing definite, then, seems to be known which would stop the empirical scientist from describing his observations in any language which he finds useful; he merely has to ensure that his definitions are not self-contradictory. A not uncommon fallacy here was exemplified by Prof. W. D. Wright who at the symposium on 'Subjective Judgments' already referred to (Bartlett, 1950) 'asserted categorically that sensation magnitudes cannot be measured, for there is no yardstick to be laid alongside of them'. The initial problem is how to define a yardstick for the first time, not how to measure along some given yardstick.

A definition of 'measurement' may therefore be something like 'the assignment of numbers to observable phenomena'. If the phenomena are consistently classified, some assignments of numbers will be more useful than others, but none will be invalid. Most formal definitions of measurement begin by some such statement. But the definitions are not left at that, for the assignment of *any* numerical symbol to *any* observable phenomenon is felt to make 'measurement' too arbitrary. Consequently, various attempts are made to restrict the meaning of measurement, usually following one or more of three lines of approach. One is to argue that only certain observable phenomena (e.g. 'quantities') are measurable, whereas others (e.g. 'qualities') are not; the second is to require that measurements should be 'physically additive'; the third is to attack the validity not of the supposedly more 'arbitrary' kinds of numerical labelling, but of the ensuing mathematics.

In the following discussion, these attempts to restrict the meaning of measurement will be queried.

II. THE SPURIOUS DISTINCTION BETWEEN QUALITY AND QUANTITY

It is customary in discussions of measurement to differentiate between 'qualities' and 'quantities'. If these terms are supposed to distinguish observable phenomena which have not (yet) been numerically described from those which have, they are acceptable tautologies for the *achievement* of measurement. If, on the other hand, the distinction is taken to be an *a priori* one—e.g. that qualities cannot be measured because they are qualities—it of course begs the question.

The most extended, and also traditional, discussion of this topic in the more scientific (as opposed to philosophical) literature seems to be that of Campbell (1928, and elsewhere). Campbell says that 'measurable properties...are those things about which certain laws are true, the laws that remove the arbitrariness from the assignment of numerals and'—

LIBRARY

here unfortunately he becomes obscure—'enable us truly to measure'. However, Campbell has to add later to this definition of measurement that 'even true measurement... is not completely free from an arbitrary element', and that in any case these 'laws of measurement' apply only to his so-called 'fundamental' magnitudes, and not to the apparently much larger class of 'derived' magnitudes. At the same time, he tells us that 'no accurate and important distinction can be drawn between the magnitudes that are measured by the fundamental method and those that are measured by the derived'. Indeed, Campbell's approach, not unlike that of other writers in this field, amounts to distinguishing not only between concepts such as qualities, quantities, magnitudes, properties, etc., generally, but also between fundamental and derived properties, measurable and unmeasurable qualities, *A*- and *B*-magnitudes, primary and secondary magnitudes and so on, without however showing that any of these classifications are logically watertight or of much demonstrable use.

If the attempts of Campbell and other such writers to reduce the arbitrary element in measurement are strangely unsuccessful, there still remains the fact that the physical scales of measurement do appear to us so very real and the converse of arbitrary. Their actual arbitrariness is, however, only hidden by long-standing conventions. Thus, Campbell's first 'law of measurement' was 'that the measurable property must have a definite order'; here the arbitrary choice of an order by some original scale-constructor is hidden by the idea that observable phenomena could 'have' an order in some absolute sense.

This matter of accepting the properties of physical measurement as if they were logical or empirical necessities should perhaps be briefly analysed here. In his recent book, *Measurement of Mind and Matter*, Scott-Blair (1949) adopts a common point of view when he says that it is clear that three dimensions are required to describe space. However, this is not only not clear, but demonstrably false. For example, all observable points, being rational, are enumerable, i.e. they can be described in *one* dimension.

Scott-Blair considers the description of houses in a symmetrical American town, with streets at right angles. Supposing, for simplicity, that there were seven streets in each direction (and that the ground is flat), the houses (or blocks) would be arranged something like the squares on a chess board. They could then be described in the usual, two-dimensional, way by pairs of co-ordinates such as:

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6	1, 7	1, 8
2, 1	2, 2	2, 3	2, 4	.	.	.	.
3, 1	3, 2	3, 3	.	.			
.	.	.	.				
.	.	.					

But one-dimensional descriptions could be made up, for example as follows:

1	2	3	4	5	6	7	8
9	10	11	12	.	.	.	.
17	18	19	.	.			
.	.	.	.				
.	.	.					

With this latter map one could still find one's way around, perform calculations on distances, and so on. The effective difference between the two kinds of mapping is in the 'principles of closeness' (as I shall call them) which are involved. Thus, in the first, any points which are 'close together' in our usually adopted sense of distance (e.g. of putting one foot in front of the other) have similar co-ordinates, in marked contrast to the one-dimensional mapping.

Now it may be true that in general practice we find those kinds of space-measurements which obey the first principle particularly useful and simple. However, in describing the location of houses in a normal town, for example, a different principle of closeness can be more advantageous, and is in fact generally used, namely that all houses facing a given street should be considered 'close together' and be given similar descriptions, with the result that there are many dimensions (street names, and numbers) and that houses which are quite close together in the usual sense (e.g. having adjoining back-gardens) may have *all* their co-ordinates quite different.

Now it seems also to be true that if we adopt the *first* principle of closeness, two dimensions are required for an appropriate space-description here, or three dimensions in general. However, as we have just seen, this principle is not a *necessary* one, and consequently the use of three dimensions for space-measurement is not necessary. There may be something deeper, i.e. less 'conventional', in the three dimensions of space (although personally I doubt it). But before the measurement of psychological phenomena is dismissed for its arbitrary conventions, some direct evidence of the lack of comparable conventions in physical measurement would seem to be required.

Without attempting to go into much further detail, there seems to be just one other aspect of 'principles of closeness' which helps to explain the apparent difference between the status of physical and that of most non-physical measurements. The various steps made in defining a system of measurement, namely the choice of the number of dimensions, the choice of ordered sequences within each dimension, and the choice of unit steps, may each be said to derive from a 'principle of closeness'. Now in non-physical scaling, *different* principles of closeness tend to be required for these various steps; in contrast, the usual physical scales all seem to derive from a single principle (namely that of 'physical additivity'). It seems to be this derivation from a *single* principle, however arbitrary it may or may not be in itself, which has given physical scales of measurement their almost overpowering elegance.

III. THE IRRELEVANCE OF ADDITIVE UNITS

The belief that measurements are 'true' or 'valid' only if the units of measurement are *additive* is very common, although there seems to be little foundation for it. Additivity here seems to mean that the *arithmetical* addition of the measurements in question must represent, or be capable of representing, certain processes of *physical* addition for the observable phenomena, or vice versa.

Campbell's writings are again usually referred to here for a formal statement of the alleged need for additivity. He embodies the requirement in the second of his 'laws of measurement'. But apart from the general lack of cogency of these laws which has already been noted above, Campbell says (1928, p. 19) of this particular law that the process of constructing a scale 'could be carried out, even if that law were not true'. Admittedly, he adds that the *usefulness* of the process depends on the truth of the law; but that is at best a much weaker proposition—and is certainly open to question—whilst pragmatically interpreted it even contradicts the assertion that the law need not in fact be true.

In general, it is surprising that it should be the issue of 'additive' or 'equal' units which most befogs the theory of measurement. Consider some set of observable phenomena already put into some ordered sequence, A, B, C, D, etc. Then it matters little whether numbers like 1, 2, 3, 4, etc. are attached to them (making each interval A–B, B–C, etc., numerically equal) or numbers such as their squares 1, 4, 9, 16, etc. (making the intervals A–B, B–C, etc., increasingly large, numerically), and so on. The numbers in the different sets are functionally related, i.e. logically they are identical. It follows that it is impossible for one set of units to be valid and another not.

Even in physics, the original choice of units is unimportant. Most physical scales of measurement (e.g. temperature, density, etc.) are, of course, in any case not directly 'additive', but even when they are (e.g. length, time, mass, etc.) the use of the measurements seldom depends on the physical additivity of the units. The distance d fallen in time t under constant acceleration g by the famous apple starting from rest, is now known to be given by the relation $d=\frac{1}{2}gt^2$. If we happened to be measuring time (as T) in non-additive log-units (which are so common in bio-physics for example) and also measuring distances in non-additive logarithmic units as D, the relation would be equally valid, being expressed as $D-2T=G$, a constant.

As for the *usefulness* of the additive units, it will be noted that the logarithmic law, being *linear*, is much simpler than the one expressed in the additive units, d and t. For example, the analysis of two sets of observations on a falling body such as:

$t=0{\cdot}5$	1	1·3	1·5
$d=4$	16	27	36

and

$t=0{\cdot}5$	9	13
$d=4$	1296	2704

would be somewhat difficult. On the other hand, measuring in logarithmic units, we could hardly help noticing that the difference between each pair of readings D and $2T$ was a constant, namely 1·2041. Equivalently, having hit on the dodge of transforming the time measurements on to a non-additive scale of the *square* of t, the simple proportionality of t^2 and d in each set of data, i.e.

$t^2=0{\cdot}25$	1	1·69	2·25
$d=4$	16	27	36

and

$t^2=0{\cdot}25$	81	169
$d=4$	1296	2704

would soon become apparent, and merely by comparing the two constants of proportionality then—16 in each case—it would be seen that the two sets of data do, in fact, obey the same law, $d=16t^2$.

The physically additive units of certain physical scales of measurement are therefore *not* always the 'simplest' to use and we seem to continue to use them (when we do) only for the sake of standardization and communicability, since they happen to be the traditional ones.

Turning now to the possible usefulness of additive units in psychology, we find that for most psychological phenomena (e.g. mental tests, subjective ratings, sensory assessments, etc.) the concept of some kind of 'physical' addition is largely irrelevant.

One could, if one wished, define a process of physical addition of intellectual functioning, for example, as the joint performance of two people on the given 'intelligence test'. But scale-construction which is merely based on additivity of *some* kind however far-fetched seems to be a waste of time. This is shown, for example, by the various attempts in psychophysics to derive sensory response scales with units 'equal' in terms of physical additions of the stimulus-variable, or by the modern work in the field of 'scale-analysis' (such as the derivation of metric scales from 'paired comparison' data, based on so-called 'assumptions' about so-called 'underlying variables' which are to behave additively): none of these scales with formally derived 'additive' units ever seems to be of much practical use.

The belief that additive units are neccessary for measurement may also be questioned on quite a different ground. It is that no reasons seem ever to have been suggested, let alone shown to be cogent, why an empirical correspondence with arithmetical addition, which is only one out of all the many mathematical operations which have been defined and studied, should be *the* criterion of 'measurability'.

IV. The use of mathematics

Instead of attempting to deny the attribute of 'measurement' to certain numerical assessments, some writers doubt the validity of most of their mathematical analysis. Stevens (1951, and earlier), for example, first adopts a very catholic definition of measurement, but considers scales of measurement as being of four different kinds. He classifies them, as either Nominal, Ordinal, Interval or Ratio scales and lists for each kind of scale certain mathematical operations which he calls 'permissible'. Stevens himself has difficulty in fitting something as crucial as *rankings* into the classification, and generally it seems unlikely that it would stand up to rigorous analysis. But the important fact to note here is the lack of evidence that the use of mathematics can be thus circumscribed.

Consider, for example, the Nominal scale, which is supposed to refer to numerical labelling without any (explicit?) choice of order or of units. Only the simplest operation may, according to Stevens, be performed on Nominal measurements, for example the counting of frequencies, but not addition, multiplication, division, etc. (The more complex functions used in applied mathematics are never explicitly 'permitted' by Stevens at all.) However, it has been found that the numbers listed in telephone directories—an instance, presumably, of Nominal measurement—cannot be used very adequately in making up tables of random numbers, and all the complicated mathematics of 'tests for randomness' had to be used to investigate the matter. The attempt to restrict mathematical analysis thus does not agree with practice.

The contrasting point of view to be urged here is that applied mathematics must do neither more nor less than describe relations between observed phenomena. For this the mode of definition of the measurements if irrelevant—for to 'measure' in Wright's sense of *using* a given yardstick, and not of *defining* one (see p. 21 above), is to adopt a routine process of describing certain observations in an agreed manner which we have merely learned to apply by rote. One could say that the validity of a mathematical analysis of measurements depends not on a correspondence of the mathematical operations with empirical *operations* such as physical addition, but on a much more general kind of correspondence, namely with empirical *relations*. It may be noted here that this principle of 'empirical validation' explains why classifications such as Stevens's and also Campbell's tend to appeal to most of us at first sight, even though on further analysis they turn out to lack relevance to our practical problems. The point is that Stevens's 'permissible mathematics' is that particular mathematics which is already empirically validated by the operations used merely in *defining* the scale of measurement. For example, suppose that the I.Q. and the weight (in pounds) of a certain child are both 110 and those of another 130. Then for the I.Q.'s we do not directly know of any meaning for the particular calculation $110+130=240$, although we may *later* find such a calculation to be meaningful. But for the weights we know at once—because of the physically additive definition of the units—that the two children put into one pan of a balance would, in fact, be balanced by a weight of $110+130=240$ lb. in the other pan.

Now, to show that the empirical validation of applied mathematics can, in fact, be sufficient, an example should perhaps be given which covers also the arguments criticized in the two previous sections, namely that the measurement of 'qualities', and addition without 'physically additive' units, are both supposedly 'illogical'.

Consider what is perhaps the most maligned mathematical operation, namely averaging. The simplest non-trivial case occurs where the observations fall into only two classes, which may be denoted or 'measured' by the numbers 0 and 1. Adding all the given measurements together and dividing by their total number gives then an average which is a proper fraction. Although no corresponding observable phenomenon or operation has been defined, this number can, of course, be interpreted, being the proportion of 1's in the data; and together with the total number of measurements, it will describe the observations completely. The process of addition and averaging is therefore meaningful, even though the measurements 0 and 1 stand for 'chalk' and 'cheese', or 'dead' and 'alive', observable phenomena which are often thought incommensurable.

V. THE STANDARDIZATION AND COMPARABILITY OF MEASUREMENTS

When people feel that something or other 'cannot really be measured', they often mean that two observations which have been given the same label do not necessarily 'mean' the same thing. A good deal of confusion arises because the latter problem has nothing to do with the applicability of numerical methods. In investigating under what conditions and in what respects two observations which have been given the same label are in fact comparable, it is irrelevant whether the label is a numerical one (a 'measurement') or not.

The problem of standardizing a new measuring technique, which is of course a fundamental one, requires for its solution nothing but a direct experimental attack (cf. Ehrenberg & Shewan, 1953). Critics of psychological measurement, and also those psychologists who would have us appalled at the task confronting them, often seem to ignore that the present-day degree of standardization and objectivity achieved in describing *physical* phenomena has only been reached after a great deal of detailed experimental work. It may now be known that the reading on certain thermometers will be about 100° C. when water relatively free from impurities is boiled at a controlled pressure; but it was not always as 'easy' as that. The work of actually investigating the conditions and respects in which a new measuring technique does give comparable readings has two aspects which require discussion here.[1]

---

[1] The unfortunate habit in the mental testing field of devising a new test, administering it to some arbitrarily chosen group of subjects, calling these 'the standardization population', and then leaving it at that, does *not* seem to call for comment.

On the experimental side there is the need for an explicit striving for standardized and 'objective' observations, and on the theoretical side there is the need for methods of analysis which facilitate the comparison of different sets of data.

First, the striving for comparable observations. It often happens in psychological work that a person has to assess or describe something the like of which he has never seen before. He therefore has to use descriptive terms or scores on whose appropriateness he cannot properly have agreed with anybody else. Clearly, no attempt can be made in such cases to obtain 'comparable' or 'standardized' or 'objective' results. Many such assessments are of course *deliberately* subjective, for example when a subject is asked to rate some picture 'on a 5-point scale for beauty'. It is irrelevant to the validity of such an assessment whether or not anyone else will agree with it. But such deliberately subjective responses do not in fact amount to the observations which the psychologist would report. Thus a subjective assessment is not reported as 'such and such a picture has a score of 4 for beauty' since others need not obtain the same result; instead, the observation is described as 'so-and-so scored the picture 4' with the implication that other qualified observers should also be able to observe so-and-so scoring 4. It follows that the occurrence of deliberately subjective assessments is no obstacle towards establishing under what conditions different observers could, in fact, observe the same results.

However, investigators in psychology, psychiatry, and in related fields often miss the point—or so it would appear—and are content for their own observations to remain on a subjective level. Thus, in many rating or sensory assessment situations, in psychiatric diagnosis, and so on, investigators tend to say what they personally consider appropriate, true, or proper, instead of trying as a primary step to achieve general agreement; but there seems to be no way (in science) of *knowing* what is appropriate, true or proper, if no one will agree. Again, even where objective techniques of observation are the avowed aim, the degree of standardization achieved is seldom investigated, let alone under constant scrutiny; but objectivity, in order to exist, must be known to exist, and be readily demonstrable.

Secondly, the need for methods of analysis which allow all experimental results to be readily compared. Rather surprisingly, perhaps, the so-called 'psychometric' methods especially developed or adapted for the analysis of psychological data are of a kind which either hinder the comparison of different experiments, or make it altogether impossible. The reason is that psychometric theory and practice deal largely in what are called 'standardized' variables. For a given set of readings, a variable is 'standardized' in this special sense if the mean of the readings is subtracted from each individual one, and if each such difference is divided by the standard deviation of all the readings. This gives a new scale of measurement which has for the given set of readings a zero mean and unit standard deviation. These properties are of course quite specific to that single set of readings; in the next experiment a *new* standardization will be carried out, and so on. The standardized scale is therefore always the same (zero mean, unit standard deviation) whatever the observations may be; or, to put it another way, the scale (in its relation to the actual observations) will differ from one experiment to the next.

Most frequently, perhaps, 'standardized' variables are used mainly by default. They are implicit, for example, when the relation between two variables is described by a correlation coefficient as follows: the correlation between variables x and y is 0·63. This merely tells us that a positive relation exists, i.e. that relative high values of x tend in some unspecified manner to go with relatively high values of y. But it does not tell us what the relation between the variables actually is: *How* does variable x vary as y varies? To be able to answer this question, we would of course need to know the means and the standard deviations of the variables as well, and it is these latter statistics which are often either ignored (by loosely talking about 'the correlation') or deliberately omitted (by explicitly working with 'standardized' variables). It follows that it is largely meaningless to compare the relationships between two variables in different sets of data by means of correlation coefficients (even when corrected for attenuation), i.e. to suppose that one has the same relationship merely because the correlation coefficients happen to be the same.

A more explicit use of 'standardized' variables occurs in what is called 'profile analysis'. Here the variables in any one study are supposed to be made 'comparable' by their standardization. But profiles from different studies are not comparable.

A more extreme case still occurs in factor analysis, where not only the *observed* variables are standardized in each analysis, but also those of the factor model itself. The complete absence in the literature[1] of valid techniques for comparing factors extracted from different analyses is therefore logically quite correct, even if not helpful.

The psychometrician's reason for using 'standardized' variables—when he makes it explicit—seems to be something like that 'he has no proper scales of measurement', or 'that he does not know what his measurements mean, anyway'. It seems certain that he can never overcome the first difficulty by using a different scale of measurement in every new experiment, or the second by making it impossible to compare the results of the different experiments. Surely the proper purpose both of numerical measurement and of applied mathematics must be to describe empirical observations in such a way as to facilitate their comparison with each other?

VI. SUMMARY

Measurement may be said to be the assignment of any numbers to any observable phenomena. It cannot be restricted to any one kind of phenomena or scale. Thus 'qualities' differ from 'quantities' merely by being phenomena to which no numbers have as yet been assigned. Again, physically additive units of measurement, which for many kinds of observation are so obviously irrelevant, are also logically unnecessary; this is shown, for example, by the fact that the choice of arithmetical addition as the mathematical function involved is a purely conventional one. Applied mathematics can be used to simplify the description of any observations, and seems to need no other justification.

The doubtful validity or usefulness of many 'measurements' arises therefore not from anything in their numerical nature, but because the underlying observational techniques either are not standardized and objective, or have at least not been shown to be so. In psychology this often follows from the use of 'psychometric' methods of analysis based on so-called 'standardized' variables, which make it impossible to compare the results of different experiments.

---

[1] One of the few explicit attempts to fill this gap seems to be in Cattell's text-book (1952, chap. 17), but it is necessarily unsuccessful. More recently Barlow & Burt (1954), for example, tell us that 'the best measure of identity (of two factors) will be the correlation between the two sets of factor-measurements'. Even if we ignore the fundamental impossibility of obtaining any factor-scores (Ehrenberg, 1953, p. 140), this proposed criterion of course suffers from all the failings of correlational analysis referred to above. Suppose, for example, that a factor extracted in one study is completely 'identical' in this sense with a factor extracted in another study, then persons with identical test-scores in the two studies will not in general have the same factor-scores.

REFERENCES

BARLOW, J. A. & BURT, C. (1954). The identification of factors from different experiments. *Brit. J. Psychol. statist. Sect.* VII, 52–6.

BARTLETT, F. C. (1950). Subjective judgements. *Nature, Lond.*, CLXVI, 984–5.

BRITISH ASSOCIATION (1938). The possibility of quantitative estimates of sensory events. (Reports of special committee). *Advanc. Sci.* p. 227.

BRITISH ASSOCIATION (1940). The possibility of quantitative estimates of sensory events. (Reports of special committee). *Advanc. Sci.* p. 332.

BURKE, C. J. (1953). Additive scales and statistics. *Psychol. Rev.* LX, 73–5.

CAMPBELL, N. R. (1928). *Measurement and Calculation.* London: Longmans, Green and Co.

CATTELL, R. B. (1952). *Factor Analysis.* New York: Harper.

COOMBS, C. H. (1950). Psychological scaling without a unit of measurement. *Psychol. Rev.* LVII, 145–58.

COOMBS, C. H. (1951). Mathematical models in psychological scaling. *J. Amer. Statist. Ass.* XLVI, 480–9.

COOMBS, C. H., RAIFFA, H. & THRALL, R. M. (1954). Some views on mathematical models and measurement theory. *Psychol. Rev.* LXI, 132–44.

EHRENBERG, A. S. C. (1953). Critical Review of Cattell's 'Factor Analysis'. *J. Genet. Psychol.* LXXXIII, 185–96.

EHRENBERG, A. S. C. & SHEWAN, J. M. (1953). The objective approach to sensory tests of food. *J. Sci. Fd Agric.* IV, 482–90.

PROCEEDINGS OF SYMPOSIUM (1953). Flavour assessment. *Chem. & Ind.* pp. 562–5.

GULLIKSEN, H. (1946). Paired comparisons and the logic of measurement. *Psychol. Rev.* LIII, 199–213.

IRWIN, J. O. (1953). Flavour assessment: A biometrician's viewpoint. *Chem. & Ind.* pp. 1348–52.

SCOTT-BLAIR, G. W. (1949). *Measurements of Mind and Matter.* London: Dobson.

STEVENS, S. S. (1951). Mathematics, measurement and psychophysics. In *Handbook of Experimental Psychology* (ed. S. S. Stevens). New York: Wiley.

(*Manuscript received* 19 *February* 1954)

A THEORY OF SHAPE RECOGNITION

I. *Introduction* (p. 30). II. *The Problem* (pp. 30–31). III. *The System* (pp. 31–36). IV. *Summary* (pp. 36–37). *References* (p. 37).

By J. A. DEUTSCH

Institute of Experimental Psychology, University of Oxford

I. Introduction

The purpose of this paper is twofold. First, it is intended to show by another practical example the fruitfulness of a view of psychological theory put forward in another paper (Deutsch, 1953). In this the author proposed a theory of motivation and learning which was of a particular logical nature. It was argued that the task of a theory in psychology is to infer the system or mechanism which is producing the phenomena of behaviour without necessarily giving the embodiment of such a system. That such a policy was practically feasible at this stage was shown by the construction of a system whose properties tallied closely with those of a motivated learning organism. Further, to show that this theory was really the description of a mechanism in general terms and not a re-description of the behaviour of an animal in highfalutin language a machine was actually built embodying the system. It is intended here to give another example of such a theory and to show how this approach is of use in the study of perception.

The second object is partially to fill what many felt to be a gap in the previous theory. The system described in the previous paper dealt with and acted on information which had already been digested. The process of classifying environmental stimulation was taken for granted. In what follows an attempt is made to elaborate a mechanism of shape perception which is complementary to the behaviour system. Though there have been previous attempts in this direction, none of the models suggested has had properties which tallied sufficiently with the known peculiarities of shape recognition in animals. Nor was it possible to put them to test as they made no particular further prediction about behaviour. Accordingly they have more the status of inventions than of explanatory theories.

II. The problem

What are these known peculiarities which any theory concerning the subject must explain?

There are six main facts:

(*a*) Animals can recognize shapes independent of their location in the visual field. It is not necessary to fixate the centre of a figure in order to recognize it nor need the eyes be moved around the contours of a figure.

(*b*) Recognition can be effected independently of the angle of inclination of a figure in the visual field. (By this is not meant tilt of the figure in depth such as occurs in shape constancy experiments, but the tilt of an image in two dimensions.)

(*c*) The size of a figure does not interfere with the identity of its shape. This of course does not hold at the extremes of size for reasons which seem sufficiently obvious.

(*d*) Mirror images appear alike. Both rats and human beings tend to confuse these. This would tend to rule out any 'template' theories of shape recognition. According to these a contour is rotated in two dimensions until it coincides with one of the many patterns already laid down. But such a superimposition cannot take place in the case of mirror images and no room is left for this particular type of confusion.

(*e*) Visually primitive organisms such as the rat and the octopus find it hard (perhaps impossible) to distinguish between squares and circles. This does not seem to be a limitation imposed by the peripheral characteristics of the optical systems. It appears to be a more central defect as these organisms can distinguish shapes which are far more alike geometrically. This type of evidence tends to cast doubt on theories which base themselves on the angular properties of figures. To quote Lashley (1938): 'The pair of figures giving greatest difficulty of discrimination where the element of acuity might play a part are the square and the circle. In several experiments these have required from 60 to 300 trials for learning and only one-fourth of the animals have shown any improvement above chance. Yet acuity is certainly not the chief limiting factor in this case.'

(*f*) These abilities, which appear to be mediated by the striate cortex, survive the removal of the major part of it. It is therefore reasonable to suppose that this ability to disentangle shape is common to all parts of the striate area and that one part of the striate area is not essential in helping the next one to operate.

This consideration would tend to rule out notions based on a scanning process. It is difficult to see how a regular scan could be maintained in the presence of extensive damage. Further, any system which requires the fixation of the centre of a figure so that it coincides with the centre of the visual field must also be ruled out. The ability is maintained even when there are extensive scotomata of central origin.

In what follows a system is proposed which will abstract from and yet satisfy the above specifications. It will also lead to fresh predictions about the discriminability of certain figures. It is therefore testable.

It is not suggested that it is the only system concerned in recognition. There must be other systems devoted to the recognition of brightness, location and size. This makes many experiments on pattern discrimination difficult to assess from the point of view of evidence for the mechanism of shape discrimination pure and simple.

III. THE SYSTEM

Let us assume that there is an array of units (or cells) arranged in two dimensions, each unit having many neighbours. This plane composed of cells has messages arriving on it from light receptors (or the retina). Each group of cells is joined to a particular retinal element in such a way that neighbouring retinal elements also excite neighbouring groups of cells on this two-dimensional array. But it is assumed that only a particular source of excitation is passed on from the retina. This is excitation from contours, the regions of the sharpest change.

Thus when a ball is focused on the retina the contour of this ball is projected on this two-dimensional array. Excitation in the form of a circle will appear on it. Thus we have the first rule.

1. Each unit on the two-dimensional array can be excited by a contour falling on the region of the retina to which it is joined.

2. When such excitation from the retina arrives each unit will pass on a pulse down what will be called a final common cable. It will also excite its neighbour. This neighbour will pass it on in its turn. Therefore when a contour is projected on the two-dimensional

array two things will occur: a message consisting of one pulse will be passed down the final common cable by each cell or unit on which that contour lies. Thus a measure of the total number of cells stimulated will be passed down.

3. Second, the contour will excite all the cells which lie next to it on the two-dimensional array. These will pass the excitation on to their neighbours but not down the final common cable. The assumption is made that a cell will pass on its excitation at right angles to the contour of which it happens to be a component. Thus for instance the excitation produced by a straight line will be a line equal and parallel to it and gradually moving away. In the case of a curve, for instance an arc of a circle, it will be also an arc of a circle moving away 'parallel' to it, but gradually decreasing as it draws away, coming to a point and then fanning out again, this time reversed (since all the advancing points of excitation move at the same speed from the point of their origin).

4. As such lateral excitation from a point in a contour advances another message will be sent down the final common cable as soon as it coincides with another point in a contour imposed on the two-dimensional array.

This message down the final common cable will at each moment give a measure of the number of points thus brought into coincidence. For example, if the figure is a rectangle, the first set of pulses sent down the final common cable will be a measure of the total number of cells excited lying on the contour (rules 1 and 2). The second set of pulses down the final common cable will be sent when the excitation travelling in a parallel line (rule 3) coincides with the two longer sides.

As these two sides are parallel to each other the excitation and the contours will coincide simultaneously and send a number of pulses proportionate to the length of these sides down the final common path all at the same instant of time. Some time later the disturbance from the two shorter lines will reach each of these lines and another message will be sent down the final common cable, indicating that a smaller number of cells have been fired off this time but also all simultaneously. Thus the message sent down the final common cable for any rectangle will consist of three sharp volleys, the proportions of the last two being governed by the ratio of the longer to the shorter sides. The message generated by a square will be two sharp volleys, both equal in magnitude. As the two pairs of sides are equidistant the advancing disturbances will hit the sides simultaneously. As they are parallel, the disturbance generated and the contours will overlap all along their length at the same instant of time, causing the cells representing the contour to fire down the final common path together.

Oddly enough a circle too would generate two sharp volleys, both equal in magnitude. All the pairs of points lying opposite each other are equidistant. Therefore the disturbance travelling across would reach all the opposite parts of the contour simultaneously—a sharp pulse equal to the first, when the circular contour first arrived on the two-dimensional array. Thus, if we agree that this is the number and relative size of the pulses that is made use of at the next stage in primitive organisms in distinguishing shapes, then circles and squares will be treated as identical. Thus the theory so far would predict the fifth finding (fact *e*) about the confusion between squares and circles. There is another group of predictions which appear *prima facie* unlikely and which have not been tested. This is that a pentagon and all odd-sided regular polygons, till there is a breakdown of acuity (central and peripheral) ought to be discriminable from a circle by these organisms where a square, a hexagon, etc., would not be.

To make the logic of this deduction clear let us consider an equilateral triangle, the simplest of the regular polygons without parallel sides. As soon as this is projected on the two-dimensional array a message will be sent down the final common cable giving a measure of the total length of contour. Immediately after this the three sides will propagate lines of disturbance parallel to themselves. As these lines of disturbance meet the contours of the triangle, there will be another message sent down the final common cable. But this time the cells on the contour will not fire off all together as the disturbance will pass across them obliquely. Their firing will be staggered according to the angle at which the disturbance and the contour interact. Thus in the case of the equilateral triangle, there will be a large initial spike or sharp volley followed immediately by a much smaller continuous discharge which will suddenly cease.

In the case of the pentagon then there will be a large and sudden first spike (as with all figures) followed by a gap (before the disturbances make contact with the opposite sides). There will then be a message of a much smaller magnitude and spread out in time, following it as the disturbance meets the opposite contours obliquely. The pentagon then, given this system of encoding and decoding ought to be quite easily discriminable from both a circle and a square, whilst these latter appear the same.

In spite of this interesting confusion it would seem that this system would derive fairly distinctive messages from most shapes which cannot be superimposed. There are two exceptions to this which are again important. This is the case of mirror images (fact *d*), another interesting property which seems accounted for. The same message will be sent down the final common cable as long as the relations of angle to length of side remain the same internal to the figure.

The second relates to figures with highly complex and irregular contours and internal lines. Unless there is some regularity of geometrical arrangement within complicated figures, the messages down the final common cable generated by these will tend to approximate with growing complexity. There would be a very large initial spike followed by a continuous signal as the many disturbances set up by contours hit other contours continuously. Any feature of equal or parallel spacing, which we normally call regularity, would set up messages which were synchronous and thus additive or sharp, giving rise to distinctive peaks in the message. Absence of these would lead to the opposite. This tallies with Lashley's (1938) experimental findings. 'The indications are that a somewhat larger number of animals fail to discriminate irregular than regular geometrical figures, but those which do discriminate require no longer to learn the one than the other type. There is evidence that those animals which learn do so by isolating some simpler cue or part figure.'

The system outlined would also display the properties of sending the same message down the final common cable whatever the inclination or position of the figure on the two-dimensional array (facts $a+b$). The message generated depends entirely on the properties of the figure itself and not on its relation with some external axes.

So far only one fact of the original six is left unexplained. This is the animal's ability to recognize shape independent of size (fact *c*). The encoding two-dimensional array will of course pass on information derived from differences in size though it discards differences in position. However, the same proportions between the components of the message passing down the final common cable will be preserved provided that the shape alone is constant. The problem of reducing messages of identical proportions to the same message or having a detector which is only sensitive to the proportions independent of size must therefore be separately considered if we adopt this scheme. A way of doing this can be envisaged, especially as the first part of the message in the system outlined above could provide an indication of the scale of the figure. The difficulty is to decide between possible alternatives as there are hardly any clues. There is a constant relationship between the initial pulse giving the length of the contour, and the size of the rest of the following message for each particular figure. This relation may be used to reduce the message for each figure to the same size, whatever the original size of this figure. If the size of the second part of the message is always divided by the size of the first, the crests of the message for a given shape will not only be of the same proportions but also of the same absolute height. (The problem of the recognition of a figure now becomes merely the problem of recognizing pulses of identical height.)

How could such a scheme be worked? For instance, the measure of the total number of cells involved signalling the length of the contour could pass on ahead. It could then act on a section of the final common cable through which the whole message would soon pass on its way to the final detector units. It would throw this section of the final common cable into a state of refractoriness proportionate to the magnitude of this measure. It would thus alter its threshold of response to the following message to differing degrees. Only a very small proportion of a message following a large pulse would survive. A large proportion of the measure after a small pulse had set the 'resistance' would pass. Thus if both these messages were derived from the same shape, let us say at different distances, they would both give rise to a message the height of whose pulses was physically identical. Hence they would pass to the final stage whither we shall not follow them. We should pass too far into the realms of speculation to give an answer to what no longer seems a problem.

So far a mechanism has been suggested to reduce a purely relational entity, a shape regardless of its location, size and inclination, to a unique set of physical events. This mechanism also shares some other rather peculiar properties with an animal's shape-recognition mechanism. It is therefore supposed that such a mechanism is in fact that used by an animal.

This mechanism has other properties which follow as predictions.

1. It should prove easier to discriminate between a square and a rectangle than between a square and a rectangle both with one side missing.

2. There should be an almost perfect transfer from an equilateral triangle to another with one side missing. This agrees with Lashley's findings, but this type of transfer should be worse in the case of other triangles, where the removal of a side should make a great difference. For instance the removal of the hypotenuse in the case of a right-angled triangle should militate more strongly against transfer.

3. Shallow curves should be undistinguishable from straight lines in the absence of other contours in the vicinity. (It may be that the factor of a change of direction in the case of the curve may provide a cue here.)

4. If shallow curves and straight lines are indistinguishable, their addition to otherwise identical figures should make these discriminable. For instance, the substitution of a curved side in a rectangle, should differentiate it from the normal rectangle. This effect should be smaller for most triangles.

5. A 'mixture' of a square and a circle should be distinct from these even though they are not distinct from each other. In fact transfer to a rectangle ought to be easier in some

circumstances, for instance, where two arcs of a circle complete a square with two opposite sides missing.

6. A scotoma of central origin when falling wholly inside a figure, where there are no contours in the vicinity, should prevent identification of the shape.

7. The prediction concerning regular polygons with an odd number of sides has already been made. It will be noted that all these predictions are independent in the sense that it would be difficult to subsume any two under a single generalization. If to the writer's greatest surprise some of these predictions receive experimental confirmation, it may be worth drawing a few more deductions from the theory.

There are, however, some difficulties.

(i) Such a mechanism could only handle one figure at a time. Consider the effect on the message down the final common cable of a square and a triangle in two different parts of the field. The two messages would be inextricably mixed up or superimposed as they would be sent down at one and the same time. They should therefore mask each other to a certain extent. But they are in fact recognized easily and separately except in one interesting condition. This is when the contours actually touch or cross. It is therefore suggested that figures are passed to the two-dimensional array on the following principle. Assume that the points in the visual field are arranged in an arbitrary order of precedence. The stimulated point which is highest in order will inhibit all the others, preventing their passage until another signal (such as might arise from the final identification apparatus) is sent up. Then the next point in order of precedence will be dealt with. Now consider the case where such points have neighbours. Assume that the effects of this inhibition are overcome by a supervening excitation passed on from this point which is inhibiting all others to those of its neighbours (if any) that were also excited, and passed on by these. (It is assumed that interrupted contours are completed.) In this way a whole contour is passed on (and other contours are kept in 'cold storage') until it is dealt with. If this was the case, masking of one figure by another would occur as soon as their outlines touched. They would now be dealt with as one complex figure. Something of the sort is borne out by everyday experience and Gottschaldt's experiment (1926). In any case this is merely to show that this property of being able to handle only one shape at a time can be circumvented by an additional hypothesis.

(ii) Higher forms can distinguish squares from circles. There is no evidence that human beings have any difficulty at all in this. It would be more plausible to propose a mechanism which would preserve a phylogenetic continuity. The present scheme cannot be made to distinguish between squares and circles as it stands. But there can be quite a simple 'evolutionary' development of it to work with greater precision and to make this distinction. The encoding part of the mechanism can be kept the same and the same message sent down the final common cable. But the message must be sent to a detecting mechanism which is sensitive not merely to absolute differences in the height of the successive pulses. A mechanism is incorporated which is sensitive to the relation between the height of the pulses and their spacing in time. These will of course depend on each other for each particular form or figure. But their proportions may vary from figure to figure, even though the proportions of the successive pulses generated by these figures may be identical. This is in fact so in the case of the square and the circle. A circle's boundary is shorter than that of a square when the distance to be travelled by the laterally spreading disturbance from contour to contour is the same. Therefore, though the two pulses sent down in each case are of the same height, there will be a greater distance between the two pulses even though they are of the same height. The analysis of the encoded information by two different means would probably improve the accuracy of recognition even where no radical confusion was possible. Such a possibility of improving the system by further development without any basic change perhaps gives the whole suggestion an enhanced biological plausibility.

(iii) It will doubtless be objected that nothing has been done to explain the lateral spread with the convenient properties postulated for it. No modus operandi has been suggested. This is admittedly a weakness of the hypothesis logically speaking. We ought in principle to state what family of machine performs this task, what interrelationships we postulate among our elements. On the other hand from the point of view of practical science to suggest possible networks of some complexity merely because they are required theoretically would be somewhat foolish, especially as there is nothing to distinguish between rival possibilities. It is really only necessary to ask whether it is plausible to assign such a property to the nervous system. Has anything like it in fact been found?

Motokawa claims to have found evidence for a laterally spreading effect on the retina, which he calls 'retinal induction'. 'For all these phenomena to take place it would be necessary that some process is propagated in a successive manner from one part to the next contiguous part.... In other words, the front of propagated induction may be supposed to proceed with a certain velocity through the retina, leaving a field of retinal induction behind it' (Motokawa, 1951). The processes of what Motokawa calls indirect induction show some of the features of the lateral spread which are postulated and which seem in many ways the least plausible physiologically. This is the ability for two types of indirect induction to cross over each other without interference with each other's spread. This property is clearly brought out in the phenomena of neutralization. Here the two processes generated by complementary colours invade each other's territory and neutralize each other. They do not form a contour as would happen if they nullified each other.

The fact that similar processes appear to take place on the retina would argue for the ease and simplicity with which the proposed system could be embodied in the central nervous system. Though no great weight is put on this analogy, it does show that the postulation of a process with somewhat similar properties in a structure of greater possible complexity is not too far fetched. It is true that what is so far known of this retinal induction indicates a greater crudity of operation than that which has been postulated for the central process. The further objection that such a central spreading process has not been suggested on anatomical grounds can be answered by saying that no-one suspected retinal induction on anatomical grounds either.

Of course until more work is done on retinal induction this possible analogy cannot be stressed too much. Perhaps the most one should say about it is that such a laterally propagated retinal process is being considered by some neurophysiologists.

IV. SUMMARY

A theory of form recognition, complementary to a previously published theory of motivation and learning is put forward. A number of facts which any theory must account for, is given. These are ability to abstract form independent of place, inclination or size,

equivalence of shape of mirror images and of squares and circles in some species, and lastly the survival of these abilities following extensive lesions of the striate area.

A mechanism which displays the requisite properties is suggested. This mechanism is also seen to have further properties which ought to manifest themselves in behaviour if the appropriate tests are made.

Lastly various objections are discussed. Aspects of the mechanism are not entirely without neurophysiological plausibility.

REFERENCES

DEUTSCH, J. A. (1953). A new type of behaviour theory. *Brit. J. Psychol.* XLIV, 304–17.

GOTTSCHALDT, (1926). Ueber den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. *Psychol. Forsch.* VIII, 261–317.

LASHLEY, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.* XVIII, 123–93.

MOTOKAWA (1951). Propagation of retinal induction. *J. Neurophysiol.* XIV, 339–51.

(*Manuscript received* 7 *May* 1954)

MOTOR AND MEMORY RESPONSES OF NEUROTICS AND NORMALS IN THE LURIA ASSOCIATION-MOTOR TECHNIQUE

By A. D. B. CLARKE

Senior Psychologist, The Manor Hospital, Epsom

I. *Introduction* (p. 38). II. *The population* (p. 38). III. *Procedure* (pp. 38–39). IV. *Results* (pp. 39–42). V. *Selective remembering* (pp. 42–43). VI. *Summary* (p. 43). *References* (p. 43).

I. Introduction

Luria's method (1932) is sufficiently well known to make a detailed explanation unnecessary; essentially, it consists of a union between verbal and motor responses in the word-association process. According to Luria, by connecting the verbal response to the motor reaction of the hand, a system can be created capable of reflecting the whole dynamic character of the central associative process; this union gives to the investigator a method by which it is possible to estimate changes in an obscure process by examination of clearly defined and easily accessible motor reactions. Thus, every sharp fluctuation and every tendency to a speech response, every marked central affective disorganization, can be observed and analysed in a compound motor reaction.

Some interesting work has been reported on the various problems of the Luria technique during the last twenty or so years, and this has been reviewed in detail elsewhere (Clarke, 1950). Suffice it to say that these experiments have shown the essential correctness of Luria's thesis. For this reason, it was decided to incorporate an association-motor experiment in a research to determine the value of objective tests as measures of neuroticism.[1] Only the main positive results of the Luria experiment are reported here; for fuller details of methodology and procedure see Clarke (*loc. cit.*).

II. The population

The subjects consisted of forty-three neurotics who were all about to be discharged from the Army on the grounds of their psychiatric disabilities and forty-three 'normal' soldiers from an Army depot. The average ages of both groups were in the early twenties and, as will be seen later, the neurotics were significantly less intelligent than the controls; this latter was probably at least in part the result of a selection factor.

III. Procedure

A word-association list containing fifty words was drawn up, a few of these being neutral; the list was as follows:

1. Boat
2. Ruler
3. Friends
4. Tired
5. Vomit
6. Reason
7. Intercourse
8. Home
9. Parade
10. Rain
11. Marriage
12. Fire
13. Mother
14. Nerves
15. Naked
16. Ink
17. Inferiority
18. Family
19. Neutral
20. Book
21. Prostitute
22. Headache
23. Me
24. Circle
25. Drugs
26. Dance
27. Castrate
28. Wife
29. Interest
30. Murder
31. Bright
32. Persecution
33. Late
34. Juice
35. Children
36. Fear
37. Lamp
38. Coat
39. Kiss
40. Grace
41. Germs
42. Father
43. Gravy
44. Nice
45. Frame
46. Leap
47. Colour
48. Venereal
49. Pain
50. Flower

---

[1] This work formed part of a research programme at the Maudsley Hospital directed by Dr H. J. Eysenck and described by him in *The Scientific Study of Personality*, 1952, London, Routledge & Kegan Paul Ltd.

The apparatus consisted of an ink-pen continuous record kymograph, an electrical 'memory drum' which presented each word in turn to the subject through a small window at 10-sec. intervals, and a pneumatic linkage for the plate on which the right hand rested.

There was one addition to this test, aimed at casting some light on the process of 'selective remembering' in relation to the list of stimulus words. There have been several investigations in the past into the concept of repression, by means of word lists being shown to the subject, and subsequent attempt by him to recall their contents. As Zeller (1950) points out in a recent review of the whole field of experimental investigation of this concept, no really significant findings emerge. Some of the writers he quotes had assumed certain words to be 'pleasantly' or 'unpleasantly' toned, thus introducing an element of subjectivity. In order to find out whether words remembered in this way are those which are pleasantly toned, and those forgotten of unpleasant affective value, it is necessary to have some objective criterion for 'unpleasant affective value'. Thus, at the conclusion of the main test the subject, without previous warning, was asked to note down in a 3-min. period all the stimulus words that he could remember. It was known that this procedure would show certain words to be more frequently remembered than others. The question thus became: Would those words which produced the greatest motor disturbance on the test be those which were least frequently remembered by the subjects? An affirmative answer would support the concept of repression or 'active forgetting' in the test situation. In addition, a second question was asked, namely: Would differences in selective remembering be apparent in the two groups, i.e. would the neurotic tend to remember different words from the controls? A final point of interest was whether there would be a significant difference in the *number* of words recalled by the neurotic and control groups; it was predicted that this measure would be positively correlated with intelligence.

Before the test commenced the subject was told that this was an experiment to see how quickly he could react, and that as each word appeared he was to say the first word that came to mind, no matter what it was, at the same time pressing down the right hand plate and keeping his left hand passively resting on a lever. Sufficient practice was given to ensure that the subject was clear about what he had to do.

IV. RESULTS

It was found that motor reaction times, failures to make motor responses, and small and large motor disturbances (measured as deviations from the straight line of the record) all showed significant point-biserial correlations with neuroticism. In addition, all these measures correlated significantly negatively with intelligence (Mill Hill Vocabulary Test). Thus, if the effect of intelligence on these scores was held constant by the part-correlation method, the correlation between the particular measure and the criterion (neuroticism) was somewhat reduced.

The best score for differentiating the two groups was found by the addition of the following items: (1) long motor reaction times, (2) right-hand 'small' disturbances, and (3) right-hand 'large' disturbances. The results gained from this composite score are shown in Table 1.

Correlation of this score with the Mill Hill Vocabulary Test gave an r of $-0{\cdot}590$. Using the motor disturbance score of 29 and above as an indicator of neuroticism, twenty-two subjects out of the eighty-six would have been misclassified (eight neurotics and fourteen control subjects).

If the effect of intelligence on this composite motor disturbance score is held constant by the part-correlation method (for this sample intelligence and neuroticism correlating to the extent of −0·320) then the correlation between the composite score and the criterion is reduced to 0·364.

The reader will have noted that all the aspects of motor disturbance on this test were negatively correlated with 'intelligence' as measured by the Mill Hill Vocabulary Test. Such a relationship seems not to have been noted in previous work, although Tendler (1945) reports it in regard to reaction times.

Table 1

	Neurotics	Controls
Mean	52·256	26·791
Variance	694·481	398·741
N	43	43

r_{pb} (neuroticism) = 0·483.
F ratio = 25·810, significant at the 0·001 level.

The Mill Hill Vocabulary Test is a test of verbal understanding; it might thus be possible that the neurotics, whose Mill Hill scores were significantly lower than those of the controls, showed disturbance as a correlate of lack of understanding. Three of the words from the stimulus list, 'Castrate', 'Inferiority' and 'Persecution', not infrequently elicited the response 'I don't know what it means'. Other words, such as 'Vomit' and 'Prostitute' were responded to in this way less frequently. One way of determining whether disturbance was partly a product of lack of understanding was to take a series of words which *a priori* and empirically would seem to have been understood by all subjects, which also showed differential disturbance on the motor level, and correlate the disturbance score for these words with the Mill Hill score to see whether a significant correlation also obtained. The following words were selected: 'Pain', 'Interest', 'Me', 'Late', 'Tired' and 'Colour'. On these words the point-biserial correlation of the composite disturbance score with neuroticism was 0·593. Correlation with the Mill Hill Test gave an r of −0·483, which should be compared with the correlation of the disturbance score on all the words of the list with the Mill Hill score: −0·590.

The fact that a significant correlation also obtains using stimulus words in which verbal understanding on the part of the subjects is likely to have been perfect, would seem to dispose of the hypothesis that the negative correlations with the verbal intelligence test were due to the difficulty of some of the words themselves. Previous investigators have viewed disturbance in association as being due primarily to affective interference, though it is true that the recent work of Albino (1948) leads him to suppose that: 'It is clear that the quantitative estimate of lability...from the graphic records shows that... (it) is not a simple quality, but is dependent on several factors.'

It will be remembered that Luria postulated a 'functional barrier' which by a regulating process kept conflict confined to cortical areas until adequate motor outlets were found. Thus the subject with the weak functional barrier showed greater motor disturbance on these tests, while the records of the subject with a strong functional barrier showed integration of response and little disturbance. Luria writes that 'in studying the genesis of the reaction movement, we see that the functional barrier does not exist in the early years of childhood but is elaborated rather late. Experimentation shows that it *appears* *about the time of the development in the child of the active organization of speech, and that it is absent in cases of serious mental retardation. All this connects the functional barrier with the higher psychological mechanisms*, and the data of aphasia *place it in direct relation to speech and symbolic mechanisms*,[1] the falling out of which produces a diffused reaction...'.[2] In relation to the above, the present findings seem to support the concept of the functional barrier, and are quite in keeping with Luria's work. What is surprising, however, is that neither he nor later workers considered the natural corollary of the above views, namely that verbal intelligence would be negatively correlated with motor disturbance (as demonstrated in the present investigation), or alternatively, the greater the intelligence, the greater the tendency towards a 'strong' functional barrier, or vice versa. This has more than academic interest, since these previous workers have treated motor disturbance as being only the effect of emotional arousal, and have neglected this further aspect.

It will have been noted that the correlation of the disturbance score with the criterion was given in both 'corrected' and 'uncorrected' forms, the former giving the criterion correlation after the effect of intelligence had been held constant. This procedure assumes that neuroticism and intelligence are orthogonal, but numerous researches, like the present one, have reported small significant negative correlations between neuroticism and intelligence. In the present study this could have been, partly at any rate, the effect of sampling. The question thus becomes, therefore, which of the two values of criterion correlations represents the 'true' value? The present research has shown that the lower the intelligence, the greater the disturbance and therefore the weaker the functional barrier; it has also shown that the lower the intelligence the greater the neuroticism. In other words, a weaker functional barrier is associated with greater neuroticism and lower intelligence. These results, showing an interdependence of these factors, are consistent with the view that lower intelligence may constitute part of the criterion, neuroticism. If this is so, then the uncorrected criterion correlation represents the 'truer' value.

In addition to the analysis of the composite disturbance score between the two groups of *subjects* (which showed a significant difference between them), it was decided to tabulate this score against the *individual words* from the stimulus list. This was designed to answer a question of some importance: Is the tendency towards disturbance on the part of neurotics and controls similar but different in degree? Response to the word-association situation has for many years been held implicitly to be strongly correlated with response to the life situations or 'areas of stress' which the particular words signify. According to the most generally accepted views of neuroticism, one would expect a significant difference in disturbance (both on the test and in the life situation), but a high correlation between the stimuli which promoted the disturbance (both in the test and in the life situation) in the two groups. Many writers, for example, have stressed the view that the neurotic personality is the one who fails to react adequately to the conflicts and difficulties which everyone experiences, but which in most people are in some sense partially resolved. In Table 2 below will be found the disturbance score on words 2–50 for each group.

The rank-difference correlation between the two sets of scores was 0·754, highly significant, showing that there is a strong tendency for the two groups to be disturbed by the same words, though as Table 1 has shown, this disturbance is different in degree in the two populations.

[2] *Loc. cit.* p. 394.

[1] Present writer's italics.

Table 2

Word	Disturbance score		Word	Disturbance score	
	Neurotics	Controls		Neurotics	Controls
Ruler	33	14	Castrate	86	64
Friends	36	18	Wife	47	21
Tired	34	8	Interest	86	40
Vomit	35	24	Murder	37	21
Reason	57	25	Bright	27	9
Intercourse	65	28	Persecution	78	50
Home	32	25	Late	52	18
Parade	36	16	Juice	38	31
Rain	22	7	Children	51	22
Marriage	42	22	Fear	65	45
Fire	34	12	Lamp	26	15
Mother	32	15	Coat	34	15
Nerves	60	35	Kiss	59	33
Naked	37	26	Grace	68	37
Ink	30	15	Germs	41	29
Inferiority	75	39	Father	43	20
Family	39	20	Gravy	41	15
Neutral	76	42	Nice	54	23
Book	28	16	Frame	34	18
Prostitute	52	34	Leap	31	12
Headache	32	21	Colour	49	15
Me	41	13	Venereal	60	33
Circle	28	19	Pain	59	15
Drugs	46	22	Flower	34	13
Dance	45	22			

V. SELECTIVE REMEMBERING

It will be recalled that at the conclusion of the test, without previous warning the subject was required to note all the stimulus words that he could remember during a 3-min. period.

The number of times each word was recalled was tabulated for neurotic and control groups. The correlation between scores thus obtained was exceedingly high, 0·957, showing that the same selective process was operating in the memory of the members of both groups (i.e. the same words tended to be remembered, and the same words forgotten, by neurotics and controls). As would be expected, the factor of recency played some part, words towards the end of the list being recalled more frequently than words at the beginning. Over and above this factor, however, the greatest tendency present was to remember words connected with family and sex (e.g. intercourse, mother, prostitute, wife, children, kiss, father, venereal, etc.). Approximately 52% of all words recalled pertain to these 'areas of stress', while only about a quarter of the stimulus words can be classified under these two headings.

The next question concerns the relation between the tendency to remember a given word, and motor disturbance on that word. In the case of the neurotics, this correlation was −0·005, and for the controls −0·082, thus showing that words which caused motor disturbance tended neither to be remembered, nor to be forgotten or repressed.

The correlation between the number of words remembered by the subjects and the Mill Hill Vocabulary Test scores was 0·549, thus confirming the prediction that the number of words recalled would be correlated with intelligence.

The controls tended to remember more words than did the neurotics (*F*-ratio: 9·764, significant at the 0·01 level). When the effect of intelligence on the 'number of words remembered' score was held constant by the part-correlation method, the negative point-biserial correlation with neuroticism decreased to insignificance. Thus the intelligence differences between the two groups account for the greater remembering capacity of the controls.

VI. SUMMARY

This paper describes part of an experiment using the Luria Association-Motor technique. The stimulus list contained fifty words, most of which were designed to have an emotional connotation. A composite score, derived from single measures which were shown to discriminate significantly between groups of neurotic and 'normal' soldiers, was found to have a point-biserial correlation of 0·483 with the criterion, neuroticism. Misclassification on the basis of this composite motor-disturbance score would have been about 25% on the population studied.

Rather high negative correlations of both single and composite motor-disturbance scores with intelligence were obtained; thus, disturbance is not only associated with affective arousals, as previous workers have supposed, but also negatively related with intelligence, in populations which approximate to normal distribution with respect to the latter. That this negative relation with an intelligence test based on knowledge of the meaning of words was not due to the subjects' difficulty in comprehending the meaning of some of the stimuli, was shown by selecting six words on which there was a significant differentiation between the groups (in terms of motor disturbance) and which both on *a priori* and empirical grounds could not have failed to have been understood. The correlation was still, however, negative and highly significant.

It has been shown that Luria's 'functional barrier', which keeps conflict in the higher cortical levels until integrated motor outlets are found, is related not only to emotional stability but also to the general mental ability and efficiency of the organism. This latter factor has not been taken into account by previous workers employing similar techniques.

Analysis of disturbance on the individual stimulus words showed a high correlation ($\rho = 0{\cdot}754$) between the two groups, indicating that the same stimuli tend to produce disturbance in both normals and neurotics, and that the difference between them is one of degree and not of kind.

The same process of selective remembering was apparent in both groups, although the controls remembered a significantly higher proportion of the stimulus words. This difference could, however, be accounted for by the difference in intellectual level between the two groups. There was no relationship between motor disturbance and the forgetting of the word, and thus the concept of repression of disturbing stimuli in this situation was not supported.

REFERENCES

ALBINO, R. C. (1948). The stable and labile personality types of Luria in clinically normal individuals. *Brit. J. Psychol.* XXXIX, 54–60.

CLARKE, A. D. B. (1950). The measurement of emotional instability by means of objective tests. Unpubl. Ph.D. Thesis, Univ. London Lib.

LURIA, A. R. (1932). *The Nature of Human Conflicts.* New York, Liveright Inc.

TENDLER, A. D. (1945). Significant features of disturbance in free association. *J. Psychol.* XX, 65–89.

ZELLER, A. F. (1950). An experimental analogue of repression. I. Historical summary. *Psychol. Bull.* XLVII, 39–51.

(*Manuscript received* 31 *March* 1954)

THE CONSTRUCTION OF AN INDEX OF INTELLIGENCE

By WILLIAM A. BELSON

Audience Research Department, B.B.C.

I. *Introduction* (pp. 44–45); (i) *The originating problem* (pp. 44–45); (ii) *General principle of the device* (p. 45). II. *Method and results* (pp. 45–49): (i) *Stage* 1: *items proposed as correlates of intelligence* (g) (pp. 45–46); (ii) *Stage* 2: *item selection* (pp. 46–48); (*a*) *General procedure* (p. 46); (*b*) *Test materials: the criterion test* (p. 47); (*c*) *Test materials: the main items* (p. 47); (*d*) *Results of item selection* (p. 48); (iii) *Stage* 3: *setting up norms* (pp. 48–49). III. *Gathering the necessary information* (pp. 49–50). IV. *Comments on the index* (pp. 50–51): (i) *Some limitations* (pp. 50–51); (ii) *Suggested improvements* (p. 51); (iii) *The principle of indices* (p. 51). V. *Summary* (p. 51). *References* (p. 52).

I. INTRODUCTION

(i) *The originating problem*

Intelligence, assessed by a measure primarily of 'g' (the Vernon Abstraction Test), had emerged in studies by the B.B.C.'s Audience Research Department as a variable highly related to a listener's ability to comprehend broadcast talks (Belson, 1952). Because of this relationship, intelligence (defined as g) became an important category for the statement of the results of such studies of comprehension.

In effect, this meant that the most meaningful statement of a programme's comprehensibility was to be made through a joint statement of (*a*) its comprehensibility to people in different intelligence groups, and (*b*) the distribution according to intelligence, of the people who eventually constitute its audience.[1]

The first of these needs could in fact be met in a fairly direct way. Comprehension tests generally involve the playing, in conference rooms, of recorded programmes (of an informative kind) to small, carefully selected groups of people and the subsequent delivery of group tests aimed at finding out if the individuals in the group have been able to grasp the various points presented in the programme. Under these conditions a formal intelligence test could be given with the other tests, so that comprehension scores could in fact be stated in terms of intelligence. It was hard to see, however, how the intelligence distribution of the programme's normal audience could be ascertained. Survey methods would have had to be used in order to meet a representative sample of the programme's actual audience, and this meant that the doorstep interview was the best testing situation that could be expected in a large number of cases—which situation was obviously quite inadequate for the delivery of the usual sort of intelligence test. If survey methods were to be used, what seemed necessary was some short assessor of g, simple to use and relatively independent of the conditions of the interview.

It also appeared at the time that it would be a decided advantage if such a device could also be used as a substitute, *in the test room*, for the formal test of g. For normal purposes there would have been little or no case for such substitution. The department's studies of comprehension are, however, usually conducted during evening sessions, half of which are devoted to listeners' discussion of their programmes, and the delivery of a formal test had been using up a considerable part of the limited time thus available for experimental work. In addition to this, the listeners who take part are generally unfamiliar with the idea of experiments or of standard testing conditions, and this meant that special preparations for the delivery of the test had normally to be made.

---

[1] An important additional function of such test results is to allow the matching, in terms of intelligence test score, of groups of people taking part in certain comprehension studies.

To meet this double need, a short device was developed—a device, that is, to make possible a rapid assessment of g under survey conditions, and usable also in the test room. This article presents the device, not so much in its departmental use as for whatever value it may have to social psychology as an independent research tool.

(ii) *General principle of the device*

Two kinds of instrument were considered, one of them a short, disguised, intelligence *test* and the other an *index* of intelligence. The principles on which they work should be carefully distinguished. In the *index*, the basic idea is that if people of different intelligence differ in respect of certain relatively stable characteristics (e.g. education, age, choice of newspaper, occupation), these characteristics may be used as indices of intelligence, the level of intelligence being estimated from them. In the short intelligence *test*, on the other hand, the subject is generally required *at the time of testing* to perform a task in such a way as to demonstrate his ability to see relationships; speed in performance would also be required. While the *short test* is no doubt less open to deliberate falsification than is the *index*, the *test* (as such) is dependent upon the vicissitudes of the interviewing situation: the nature of the training of the interviewer is likely to be important enough, but in addition to this, it is necessary to think seriously of how typical are the intellective performances of people 'caught on the doorstep' as it were; then again there is the difficulty of applying any effective control to the time allowed for the test. The index approach, on the other hand, appears to be relatively independent of the conditions of the interview and does not involve the question of timing. Accordingly, research was directed mainly towards the development of an index.

II. METHOD AND RESULTS

There were three stages in the construction of the index: (1) the compilation of a list of characteristics or items which might with some plausibility be thought to be correlated with intelligence (g); (2) the testing of these items against a criterion of intelligence (as defined) in order to select that combination of them yielding the highest multiple-correlation with the criterion; (3) the norming of the index (consisting of the selected items) for the general adult population of Great Britain.

(i) *Stage* 1: *items proposed as correlates of intelligence* (g)

The listing of proposed items was restricted to some extent by the requirement that the final index should fit into the general context of one or another of the department's questionnaires and also, to a slight degree, by an emphasis on items likely to be useful in themselves. Those finally listed for study included: occupational level, educational level achieved, age, choice of newspaper, frequency of book reading, frequency of cinema-going, degree of crowding in the home, hobbies and interests, attitude towards the criticism of one's own beliefs (by broadcasters), degree of attachment to radio, attitude towards listening to the Third Programme, active membership of a society or club, the buying of Penguin or Pelican books, use of the Radio Times in the selection of programmes, general opinion of B.B.C. programmes, likes and dislikes in relation to specific programmes, programme service preference (that is, preference for the Home Service, or the Light Programme, or the Third Programme).

It was of course clear at the outset that occupational level is related to intelligence (Foulds & Raven, 1948; Lingwood, 1952; Vernon, 1949). To make this item apply to the whole adult population, it was proposed to rate housewives by job before marriage and retired people by last job held, though it was realized that this would probably add to the predictive inaccuracy occasioned by the reported wide range of intelligence within any one occupational group. While it also seemed reasonable to expect a useful degree of correlation between educational level and intelligence, the majority of the adult population in Great Britain did not proceed beyond elementary school (leaving at 14 or 15 or earlier) and this suggested that educational level might not in fact be a very discriminating predictor. The inclusion of age in the list had its support in the various studies already made of the relation between age and intelligence test score (Jones & Conrad, 1933; Miles & Miles, 1932; Vincent, 1952), though clearly much would depend upon the nature of the intelligence test (Lorge, 1936; Wechsler, 1950).

(ii) *Stage* 2: *item selection*

This stage involved: (1) determining the extent of relation to intelligence test score of each of the proposed items; (2) determining, on the bases of these correlations, which combination of the proposed items would give the most accurate prediction of intelligence test score; (3) determining the degree of accuracy to be expected of this prediction.

(*a*) *General procedure.* Some 270 people, constituting a sample representative of the adult population of Great Britain in terms of occupational level, age, and sex, took part in this stage of the work. They were all drawn, by written invitation, from those people who were interviewed previously in one or another of the department's daily surveys of listening and viewing[1] and they met on various evenings in London conference rooms in groups of 15 to 20. While part of the evening was devoted to explaining the requirements of the study and to winning co-operation, the essential steps were: (i) the gathering of information allowing classification of each person in terms of each of the characteristics or items to be examined; (ii) the delivery of a criterion test of intelligence (g). All the work was of a paper and pencil kind. Details of the main items and of the criterion test are given below under *Test materials*.

Previous research had established the rates at which people of different age, sex and occupational level would turn up in response to invitation. These differential rates were taken into account in the issuing of invitations but, even so, a group secured by this method of invitation could only approximately fulfil sample requirements; once the group was in the test-room, however, the necessary details (about occupational level, age and sex) were gathered and this made possible the subsequent refinement of the sample so that it finally approximated very closely indeed to the general adult population (in terms of occupational level, age and sex).

---

[1] These are quota surveys, involving 2500 interviews each day. The controls are social class, age, sex and whether or not gainfully employed.

(*b*) *Test materials: the criterion test.* The test used in this study as a criterion of intelligence was the Vernon Abstraction Test. This test, which consists of forty abstraction-type items (Vernon & Parry, 1949) and imposes a time limit of 15 minutes, has only small verbal and numerical loadings and is primarily a test of g. Though it is a paper and pencil test, very little writing indeed is involved in it. It was this test which, as a measure of g used in earlier research (Belson, 1952), had brought out the importance of the relation between comprehension and g. This fact was one reason for its choice as a criterion in the present study. Another reason, discussed more fully later on, was that its use here made it possible to draw directly from a large volume of data gathered in earlier research.

(*c*) *Test materials: the main items.* Of the twenty-one items proposed as components of the index, only occupational level, age, educational level and 'frequency of book reading' are described in any detail here, as these were in fact the four items yielding the highest multiple-correlation with the criterion (details of the other items, along with their correlations with the criterion are available from D.S.I.R. Technical Information and Documents Unit).

The scaling of occupational level required the collection of details of the subjects' occupational background. Thus each of the subjects who was in a job completed a brief occupational questionnaire giving the name of his or her job, a short statement (of not more than two lines) of what was involved in the job, qualifications (if any) required for the job, and the number of people (if any) working under his or her authority. Housewives gave the same information for job before marriage and retired people referred to last full-time job before retirement. This information made it possible later on to place each person in one or another of seven occupational groups scaled according to 'skill and training required for the job'. These groups were: higher professional and administrative; lower professional, technical and executive; highly skilled; skilled; moderately skilled; semi-skilled; unskilled. On this scale each group is described in terms of skill and training required and is illustrated by sample occupations. The construction of the scale, which had its beginnings in one presented by Spielman & Burt (1926), was influenced by research data giving mean intelligence test scores in various occupations (Foulds & Raven, 1948; Lingwood, 1952; Vernon, 1949).

For the scaling of educational level, the following educational information was recorded: age of ceasing full-time education, type of school attended, the nature of any (further) education completed since leaving school, any examinations passed or certificates gained. This information then made it possible to place each person in one or another of eight educational groups scaled according to 'general (or academic) educational ability or equipment'. (Both the occupational and the educational scales are available from the D.S.I.R. Technical Information and Documents Unit.)

Age was recorded directly on a personal detail form. Classification according to 'frequency of book reading' was based on the answer to the question 'have you managed to read any books recently?', the subject answering 'yes', 'no' or perhaps giving no reply at all; this item is a very close approximation to one used successfully by Goossen in his American study (Goossen, 1950). With the exception of 'choice of newspaper' (where the names of newspapers read were given), interests and hobbies (where a list of these was volunteered) and one or two other items, the classifications were based on multiple-choice responses to questions of this kind.

(*d*) *Results of item selection.* In the final analysis the more highly predictive of the individual items were occupational level (+0·65), educational level (+0·57), age (−0·48), book reading (+0·38) and 'attachment to radio' (+0·35).[1] These correlations and the inter-item correlations are shown in Table 1. Occupational level and age yielded a multiple-correlation with intelligence of +0·785, educational level increasing it to 0·798 and 'book reading' increasing it to 0·808. It should be noted that while educational level is the second most powerful *individual* predictor, it is quite highly correlated with occupational level; age, on the other hand, has practically zero (−0·07) correlation with occupational level. Accordingly it is not surprising that, for the purpose of maximizing correlation with the criterion, age is a better item to use in conjunction with occupational level than is educational level.

Table 1. *Criterion and inter-item correlations*

	Intelligence (*g*)	Occupation	Education	Age	Book reading
Occupation	+0·65	—	—	—	—
Education	+0·57	+0·64	—	—	—
Age	−0·48	−0·07	−0·15	—	—
Book reading	+0·38	+0·28	+0·23	−0·14	—
Attachment to radio	+0·35	+0·38	+0·37	−0·27	+0·53

From a practical point of view, there seemed much to be gained by limiting the index to occupational level and age and sacrificing the small gain in validity coming with the use of the third and the fourth items for the simplicity of a two-item index. Accordingly, only occupational level and age were retained.

(iii) *Stage* 3: *setting up norms*

Norming was carried out in two steps. In the first step, the test results of an approximately rectangular sample of people (that is, a group with an approximately equal number of people in each relevant subsection) was used to establish the mean intelligence test scores in each of the various occupation/age groups (e.g. skilled, 26–35 years; unskilled, 66+years). The bulk of the material for doing this was fortunately available from an earlier study in which each of over 1000 people of widely different age, occupational level and intelligence, completed both the intelligence test and the occupation/age detail form used in the present study (Belson, 1952). There were six age groups at each of six occupational levels, making thirty-six occupation/age subgroups in all. The number of people in each of these averaged about 30. Mean intelligence test scores of the six subgroups at each occupational level were plotted and a limited amount of smoothing was done. The thirty-six adjusted mean intelligence test scores are set out in Table 2.

Secondly, appropriate questions about occupational level and age were included in a random survey of the adult civilian population of Great Britain (approximately 4700 interviews) being carried out at the time, thus making it possible to determine the percentage of the population falling into each of the thirty-six occupation/age groups. These percentages appear in Table 2.

---

[1] Those people answering 'yes' to the question about book reading tended to be more intelligent than those answering 'no' or giving no reply; those using extreme statements on the attachment scale or giving no reply tended to be less intelligent than those choosing the more moderate statements.

Since the percentage size and the mean test score of each group was known, it was then possible, through a simple cumulative frequency procedure, to determine which occupation/age groups fell into the most intelligent 10% of the population, which fell into the next 10% group, which into the third 10% group, and so on right down to the lowest scoring 10% group.

Table 2. *Percentage size of each occupation/age group, mean test scores and decile groups*

Age group	Occupational level* 1/2	3	4	5	6	7	Others	Percentage adult population of Great Britain† (excluding servicemen)
21–25 years								
Mean test score	34·9	33·8	29·2	23·5	21·0	18·0	—	—
% G.B. population	0·1	0·8	1·5	2·0	1·2	0·7	0·2	6·5
Decile group	I	I	II	III	IV	VI	—	—
26–35 years								
Mean test score	33·8	31·8	27·4	22·4	19·5	16·0	—	—
% G.B. population	0·5	1·9	5·7	4·5	4·3	1·9	0·6	19·4
Decile group	I	I	II	IV	V	VII	—	—
36–45 years								
Mean test score	33·4	31·3	24·8	19·8	17·3	16·0	—	—
% G.B. population	0·9	2·7	5·0	6·5	4·8	3·0	1·0	23·9
Decile group	I	I	III	V	VI	VII	—	—
46–55 years								
Mean test score	30·7	29·6	22·2	17·3	15·5	13·8	—	—
% G.B. population	0·8	2·0	5·3	5·2	4·5	3·2	1·1	22·1
Decile group	I	I	IV	VII	VIII	IX	—	—
56–65 years								
Mean test score	28·0	27·4	18·8	13·8	12·5	10·2	—	—
% G.B. population	0·8	1·1	2·4	3·3	3·1	2·8	1·1	14·6
Decile group	II	II	VI	VIII	IX	IX	—	—
66 + years								
Mean test score	22·7	21·5	13·0	10·0	9·0	7·5	—	—
% G.B. population	0·6	0·8	2·3	3·6	2·4	2·2	1·6	13·5
Decile group	III	IV	IX	X	X	X	—	—
All ages (totals)								
% G.B. population	3·7	9·3	22·2	25·1	20·3	13·8	5·6	100·0

* 1/2 = higher and lower professional; 3 = highly skilled; 4 = skilled; 5 = moderately skilled; 6 = semi-skilled; 7 = unskilled; others = persons without occupational record or unclassifiable.

† Based on the random sample referred to on p. 48.

The ten groups thus established were numbered, I, II, III, ..., X and were called the first, second, third, etc., decile groups. The decile groups into which the various occupation/age combinations fell are entered in Table 2 in Roman type and this set of entries may be used without further reference to any of the other entries to estimate intelligence level from occupational level and age. (In normal practice Table 2 would be reduced to a simple two-way table with only decile groups entered in the cells.) From Table 2 it will be seen that if a person interviewed in a survey is of occupational group 5 and of age group 46–55 he will be rated as falling into the seventh decile group (VII) in respect of intelligence.

III. Gathering the necessary information

In gathering (during a survey interview) the information required for use of the index, questions about age and occupation may be put towards the end of the interview. Age should be asked for directly and questions about occupation should be as follows.

(*a*) What is your job? ....

If *housewife* or *retired*, ask:

What was your job before you married*/retired? ....

....

If informant names more than one job, record the last one.

(*b*) What kind of work do/did you do in this job? ....

....

In some cases (e.g. bus conductor, railway porter) question (*b*) will not be necessary.

(*c*) Do/did you happen to have anyone working under your authority in your job?...........................

If 'yes', ask:

How many?...

(*d*) What training or qualification did you need for the job? ....

* If housewife and no previous occupation, refer all these questions to husband's job.

The interviewer merely records the information without making either rating or judgement. Ratings are made by office staff trained in the use of the scale, and of course, the more extensive their knowledge about jobs, the better their ratings are likely to be. In test room sessions the same information can be collected quite easily through a standard form.

IV. COMMENTS ON THE INDEX

(i) *Some limitations*

It is important to point out, first of all, that this index was designed as a classifying device for the *grouping of results*, and that though it works *through* prediction of individual intelligence, it is not meant for the prediction of *individual* intelligence as such. Though it might be of some use as a general screening device, it is essentially a tool for sociological inquiries involving the grouping of results.

Secondly, it should be pointed out that the multiple-correlation of $+0{\cdot}785$ is based upon information gathered under test room conditions and not in doorstep interviews, even though the index is required to be valid under such doorstep conditions. Had the items selected been more of a 'test' nature this might have been crucial, but with variables like occupation and age it was tentatively assumed that there would not be appreciable differences in the information collected under the two conditions. Further research on the point should not, however, be ruled out. Somewhat more serious, perhaps, is the limitation arising out of basing the index on the responses of a London sample, for there is evidence (Vernon, 1947) to suggest that some regional differences, small but significant, may be expected in mean intelligence test scores within a given occupation. If this is so, it means that in national surveys some regional adjustment is necessary.

Thirdly, there is no point in using the present index, viewed as a general research tool, unless *g* is in fact related in an important or relevant way to the variable under study. Its use in studies of comprehension (of broadcast talks), for instance, arose out of the high degree of correlation found to exist between *g* and comprehension score, a relation probably accentuated by the essentially non-stop nature of the radio talk. It is worth noting in this connexion, however, that *g* is not necessarily the most powerful single predictor of comprehension available. Conceivably an intelligence test with an appreciable verbal loading and delivered at speed would correlate with comprehension score at a slightly higher level than does a test primarily of *g*. Unfortunately, a test of this other kind, really suitable for the very special test-room conditions concerned, was not available at the time of the collection of the bulk of the data used in this project—namely the set of records giving age, occupational level and a measure of g referred to above (Belson, 1952).

(ii) *Suggested improvements*

No doubt a number of improvements could be made in the index. In the first place it should be possible to continue with the adjustment of the scale in accordance with research findings which give mean intelligence test scores in specific occupations. How far this could be carried without destroying the basic dimension of 'skill and training required' is of course a matter of some importance, and in the end depends upon the relation between that variable and intelligence. The trend in such an adjustment is clearly towards a replacement of the 'skill and training' scale by a direct intelligence scale, each section being copiously illustrated by occupations for which the mean intelligence test score has been determined by research.

Another improvement might be made, perhaps, through using educational information in the interpretation of the occupational details provided; again, one or two questions about occupation might be added (e.g. whether monthly or weekly paid) to increase discrimination between jobs otherwise similar in description; and of course the better the fund of general occupational information, the more accurate will be the placement of specific occupations upon the scale. Again, additional items (i.e. items additional to the two actually used) might be added to the index, though there is probably only a narrow margin within which this process could in practice be continued.

(iii) *The principle of indices*

Quite apart from the value which this index of g may have as an independent research tool, it does seem possible to employ the principles used in its construction to achieve rapid assessment under survey conditions (and at a level of accuracy suitable for the grouping of results) of a certain range of other abilities and variables, the measurement of which normally requires the administration of relatively lengthy tests under carefully controlled conditions.

V. SUMMARY

For certain purposes relating to audience research it was found necessary to develop a simple method of estimating intelligence which would be independent of the administration of conventional tests. To this end an 'index' consisting of grades on a ten-point (decile) scale was evolved. The index in its final form was based upon two items, 'age' and 'occupational level', which together yielded a multiple correlation of +0·785 with intelligence (defined as g and assessed by the Vernon Abstraction Test). Occupational level, which is here used in a specialized sense, refers either to present job or job before retirement or, for housewives, to job before marriage. A simple two-way table was developed for the direct conversion of age and occupational level into an intelligence grade. The index as constructed and normed is intended for use in Great Britain.

REFERENCES

BELSON, W. A. (1952). The comprehensibility of 'Topic for Tonight'. An Audience Research Department (B.B.C.) report, LR/52/1080.

FOULDS, G. A. & RAVEN, J. C. (1948). Intellectual ability and occupational grade. *Occup. Psychol.* XXII, 197.

GOOSSEN, C. V. (1950). The Goossen hidden intelligence test. *Pub. Op. Q.* XIV, 759.

JONES, H. E. & CONRAD, H. S. (1933). The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monog.* XIII, 223.

LINGWOOD, J. (1952). Test performances of A.T.S. recruits from certain civilian occupations. *Occup. Psychol.* XXVI, 35.

LORGE, I. (1936). The influence of the test upon the nature of mental decline as a function of age. *J. Educ. Psychol.* XXVII, 101.

MILES, C. C. & MILES, W. R. (1932). The correlation of intelligence scores and chronological age from early to late maturity. *Amer. J. Psychol.* XLIV, 44.

SPIELMAN, W. & BURT. C. (1926). The estimation of intelligence in vocational guidance. In *A Study in Vocational Guidance*. Industrial Fatigue Research Board Rep. No. 33.

VERNON, P. E. (1947). The variations of intelligence with occupation, age and locality. *Brit. J. Psychol. Statist. Sect.* I, 52.

VERNON, P. E. (1949). Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.* XXIII, 58.

VERNON, P. E. & PARRY, J. B. (1949). *Personnel Selection in the British Forces*. London: University of London Press.

VINCENT, D. F. (1952). The linear relationship between age and score of adults in intelligence tests. *Occup. Psychol.* XXVI, 243.

WECHSLER, D. (1950). *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins Co.

(*Manuscript received* 29 *April* 1954)

SPEECH-BREATHING ACTIVITY—A MEASURE OF TENSION AND AFFECT DURING INTERVIEWS

By FRIEDA GOLDMAN-EISLER

Medical Research Council Unit for Research in Occupational Adaptation,
Maudsley Hospital, London

I. *Introduction* (pp. 53–54). II. *The problem* (p. 54). III. *The measures* (pp. 54–56): (i) *Method of recording breath rate during speech* (p. 54); (ii) *Measuring speech production* (pp. 54–56). IV. *Consistency and variability of measures* (pp. 57–60): (i) *Material* (p. 57); (ii) *Results* (pp. 57–60). V. *Discussion* (pp. 60–62): (i) *Psychological significance of breath rate* (pp. 60–61); (ii) *Psychological significance of the number of syllables per expiration* (pp. 61–62). VI. *Summary* (p. 62). *References* (p. 63).

I. Introduction

Interviewers of all kinds, especially psychotherapists and psychiatrists, have for long been beset by the problem of how to assess the fluctuations of tension and affect which take place in persons while being interviewed. It is increasingly being realized (Mowrer, 1953; Ruesch & Prestwood, 1949) that the best way to study such changes is through indicators of mental and emotional states measured during, and by way of, the speech process itself. The present study is meant to contribute to the solution of this problem.

In a search for such indicators the writer (1954*a*) studied time sequences, rhythm, and rate of speech, as well as a number of grammatical categories in language obtained through time and speech recordings. These measures were tested for consistency within, and sensitivity in discriminating between individuals, as well as between situations (Goldman-Eisler, 1952, 1954*a*). The speech rate, or speed of talking, was found to be the most consistent and most sensitive of these measures, and variations in the rate of speech were found to be indicative of a total change in linguistic activity. The changes, which included content and affected both speakers, were of a compensatory and self-regulating nature. This finding suggested that variations in the speed of talking may serve to restore the balance of the linguistic equilibrium where this has been disturbed as a result either of strong affect or of tension having been aroused during the interview. As a consequence of these studies the rate of speech production was subjected to further detailed analysis (Goldman-Eisler, 1954*b*).

So far the following facts concerning speech rate have emerged. Speech rate is relatively constant for any person from one conversation to another when calculated as an average over the total length of an individual's speech output. When speech rates were calculated for the individual utterances in a conversation (as the average number of syllables uttered each minute) consistency was confined to long utterances containing 100 syllables or more; for shorter utterances a person's speed varied over five times as wide a range as for long utterances. Moreover, short utterances were, on the average, faster than long utterances. The fast rates of speech which were common in short utterances rarely occurred in utterances more than 60 syllables long, and never in utterances of more than 100 syllables. The greater speed of short utterances was assumed to be due to freedom from inhibiting factors, both physiological (need to take breath) and psychological (emotional blocking or need to think), which are bound to slow down speech to some extent. The psychological implications of speed and acceleration in speech have been discussed elsewhere (Goldman-Eisler, 1954*b*).

II. THE PROBLEM

The present paper deals with the attempt to break down the measure of speech rate into constituent measures derived from respiratory behaviour during speech. In view of the importance of breathing to the production of speech on the one hand, and of the emotional significance of respiratory variations on the other, measures of speech-breathing activity promised to be more immediately relevant to the changing states of tension and affect during interviews than speech rate unqualified by further definition could be expected to be.

The problem to be solved was twofold. A workable method of measuring breathing during speech had to be found, and such measures of speech-breathing activity had to be developed as would prove psychologically relevant. The writer's object in the present paper is primarily to call attention to the technique.

III. THE MEASURES

(i) *Method of recording breath rate during speech*

The recording of breathing by the usual apparatus, pneumograph, spirometer, etc., seems most undesirable if the speakers are to remain as uninhibited and spontaneous as is necessary if we want to make investigations about natural and spontaneous speech, and use our type of analysis in psychiatric interviews and psychotherapy.

The writer discovered that by listening carefully she could hear quite distinctly the noise produced by speakers when inhaling air. Two observers counting inhalations of recordings, for which verbatim manuscripts were at hand, were practically in complete agreement. The reliability of this method, of course, presupposes a highly sensitive microphone and most conscientious and accurate listening.

(ii) *Measuring speech production*

Using the above method the following measures for each utterance are at our disposal:

(1) Duration of utterance, t.

(2) Number of syllables in each utterance, N_s.

(3) Number of inspirations in each utterance, N_i.

From these we derive:

(4) The rate of respiration, i.e. the number of inspirations per minute, $N_i/t = RR$ (or rr for low rate of respiration).

(5) The rate of speech, i.e. the number of syllables per minute calculated for each utterance, $N_s/t = SR$.

(6) Output per breath, i.e. the number of syllables per expiration (or rate of expulsion of syllables), $N_s/N_i = ER$ (or er for low expulsion rate).

(7) Its reciprocal, measuring the proportion of the returned air current per syllable spoken expressed as a percentage, $N_i/N_s \times 100$ ($V\,\%$ = ventilation percentage).

From the above it may be seen that the rate of speech N_s/t is the product of breath rate and output per breath, i.e. $N_i/t \times N_s/N_i = N_s/t$. From this formula it follows that there could be an infinite variety of combinations between the two factors N_i/t and N_s/N_i possible for any particular rate of speech, the number of syllables per expiration decreasing as breath rate increases and vice versa. Each particular rate of speech is a hyperbola, the two sides being breath rate and number of syllables per breath. Thus when a person speaks at a rate, e.g., of 300 syllables per minute, he may have achieved it by inhaling 10 times and uttering an average of 30 syllables between each inspiration, or 30 times speaking only 10 syllables between each breath, or anything between as well as beyond these figures.

Thus the same rate of speech would cover very different physiological and psychological processes operating in the production of it. This is reminiscent of the different moods which may be covered by the same metronomic tempo in music.

We are proposing to describe any utterance in respect of the two processes which combined give us a measure of the speech rate, namely the respiratory activity and the degree of its utilization as speech sounds. The former is given in the rate of respiration (number of inspirations per minute) and the latter as the number of syllables per expiration being, as will be shown later, also a measure of the rate of escape or expulsion of the returning (expired) air current. Fig. 1 with the speed hyperbolas (on double-logarithmic paper they are straight lines) drawn in allows us to locate each utterance provided we know the number of syllables, the number of inspirations contained in it, as well as its duration. Examples of graphical representation on double-logarithmic paper are on p. 56, and we can see that even within the narrow confines of constant speed, utterances differ widely in respect of breathing activity (measured by RR) and the economy of distributing it per syllable. Speed of talking and rate of respiration are independent dimensions of speech, at least in general, although on occasions we meet with speech situations in which acceleration was achieved by one kind of mechanism chiefly, either by increasing or decreasing the rate of respiration. Both cases are possible, and we may expect each to have different psychological connotations.

Thus utterances can be plotted as is illustrated in Figs. 1 and 2.

The scheme in Fig. 1 represents the various ways of accelerating utterances. In the left bottom corner of the rectangle ($rr \times er$) we would place utterances produced with minimal supply of air and minimal output per expiration; they would thus be utterances of minimal speed. Moving vertically upwards we would find utterances with increasing output of syllables per expiration carrying minimal ventilation, while the bottom horizontal side of the quadrangle would be for utterances of increasing air supply and minimal output of syllables per expiration. As utterances move from there along the right vertical towards the right top corner ($RR \times ER$) we have reached the maximum speed achieved by maximum breathing activity, and maximum air expenditure. If for this scheme we use double-logarithmic paper and plot each utterance according to respiration rate (abscissa) and output of syllables per expiration (ordinate) the rate of speech will be the diagonal while the two ordinates RR and $N_s/N_i = ER$ are two sides, according to the above formula.

Such graphical presentation of utterances in any one conversation illustrates at first sight the basic differences in the nature of speech production even where there are no differences of speech rate (Fig. 2).

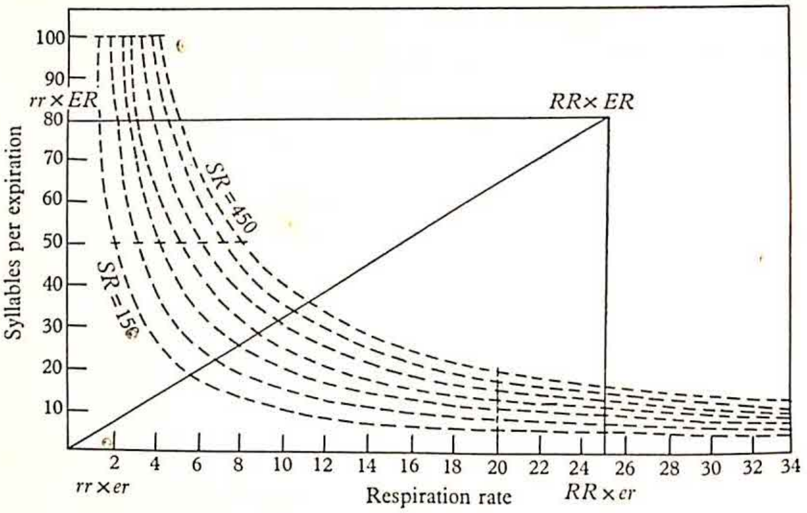


Fig. 1. The broken lines indicate the speech rates (*SR*) rising by 50 from 150 to 450 syllables per minute achieved at respiration rates up to 34 a minute and an output of up to 100 syllables per expiration.

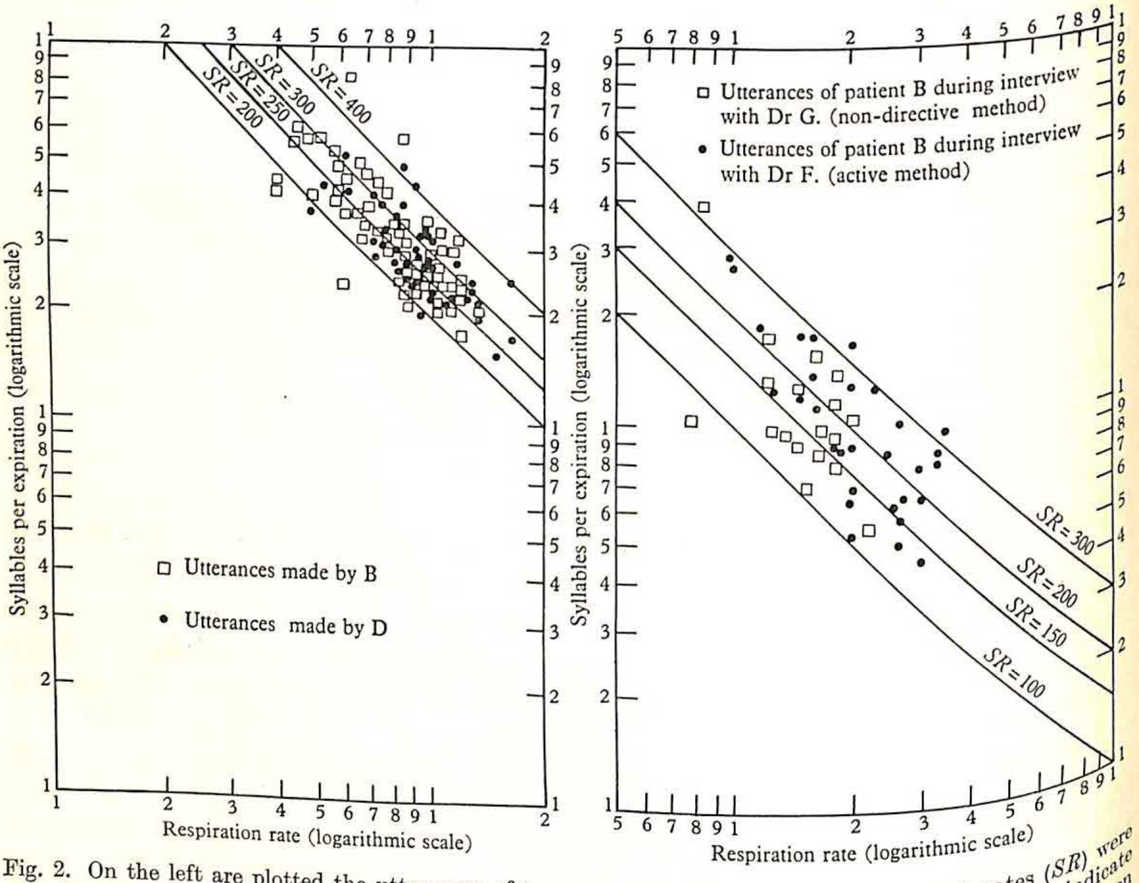


Fig. 2. On the left are plotted the utterances of two persons in conversation whose speech rates (*SR*) were about equal, but who differed in output per expiration and respiration rate. The logarithmic scales indicate respiration rates of 1–20 on the left, and 5–100 on the right, and for syllables per expiration of 1–100 on both sides. On the right are plotted utterances from two interviews of a schizophrenic patient B with two different psychiatrists. Dr G. used the non-directive, and Dr F. the active technique of interviewing. The two interviews are differentiated not so much by speech rates as by the higher degree of ventilation with which B responded in the active interview.

IV. CONSISTENCY AND VARIABILITY OF MEASURES

The above observations and deductions provide information as to the possible relations in speech between breath rate, number of syllables per expiration, and speech rate. The individual examples illustrated how these possible combinations may appear in actual situations in some individuals.

The 'goodness' or usefulness of the measures introduced depends on their capacity for measuring individual differences.

Tests of consistency and discriminative capacity of the rate of respiration in speech, as well as of the number of syllables per expiration (or its reciprocal), carried out by means of the statistical method of analysis of variance, led to the following results.

(i) *Material*

The material consisted of twelve discussions between four adults of superior intelligence in different combinations, seven interviews between the experimenter and four young female neurotics, one schizophrenic, and three research workers on the staff, as well as four diagnostic interviews by a psychiatrist with patients attending the outpatient clinic. Altogether, twenty-three interviews containing 829 utterances served as the material for this investigation.

The number of syllables, as well as of inspirations, were counted and speech and breath rate calculated, as well as the output of speech (number of syllables) per breath and the reciprocal of this value, the ventilation percentage calculated.

(ii) *Results*

Two Analyses of Variance were carried out. The one was based on the four individuals who had discussed controversial subjects with each other. Three recorded discussions were available from each of them, and it was therefore possible to test their individual consistency for different conversations with different interlocutors, on different occasions, and discussing different topics, and to compare the variance within each of these individuals with the variance of the tested measurements from person to person. This group was a comparatively homogeneous one consisting of research workers and academic personnel. The second Analysis of Variance was based on utterances derived from a mixed group of twelve individuals, patients and normals, from whom only one interview-recording was obtained. The internal consistency for these measurements was obtained by comparing those taken in the first half with those from the second half of the interview.

Tables 1–4 show the means and standard deviations for these two groups of respiration rate and of the degree of ventilation as measured by the reciprocal of the number of syllables per expiration. (The reciprocal was chosen for statistical calculation, being more normally distributed.)

These data were obtained by two interviewers; the table of means shows that the differences cannot be due to the interviewer's influence and that the subjects interviewed by the experimenter (*Ma*, *Nu*, *Stn*, *S.O.*, *S.* I, *S.* II) differ much more among themselves (this group includes patients as well as normals) than the subjects interviewed by *B* who are all patients. Neither are systematic differences between the patients and the normals evident, but the groups are too small to answer the question of such differences.

It is clear from these tables that not only the rate of respiration but also the degree of ventilation per speech unit, or reciprocally the number of syllables spoken per expiration, are measures of high consistency and capacity of discriminating between individuals. At the same time they are sensitive to fluctuations in speech-breathing behaviour, as the significant differences between the first and second halves of the interviews in the second set of measurements (Tables 3 and 4) show.

Table 1. *Means and standard deviations of respiration rates of four intelligent normal adults based on* 463 *utterances made in twelve conversations* (*three per person*)

	Persons											
	A			*B*			*C*			*D*		
Conversations	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*
1	6·06	2·59	36	8·42	2·63	43	8·81	5·02	59	8·48	2·43	59
2	5·66	2·31	13	6·25	2·50	17	9·16	4·66	42	9·76	2·65	48
3	7·63	3·97	29	8·32	3·62	24	9·95	3·56	31	8·27	3·20	62

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Persons	3	430·89	143·63
Conversations	8	10·24	1·28
Error	451	5602·86	12·13
Total	462		

F, persons/error = 11·84. *P* at 0·001 = 5·42 (D.F. 3∞). ϵ* = 0·957.
F, persons/conversations = 112·21. *P* at 0·001 = 15·83 (D.F. 3 and 8).

* Epsilon is a measure of correlation based on the analysis of variance (see Peters and Van Voorhis, *Statistical Procedures and their Mathematical Bases*, 1940, pp. 325, 353. New York: McGraw-Hill).

Table 2. *Means and standard deviations of ventilation percentages of four intelligent normal adults based on* 463 *utterances made in twelve conversations* (*three per person*)

	Persons											
	A			*B*			*C*			*D*		
Conversations	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*
1	2·90	1·13	36	3·48	1·18	43	5·14	3·23	59	3·19	0·98	59
2	2·28	1·06	13	3·54	1·84	17	5·04	2·47	42	3·59	1·00	48
3	3·08	1·39	29	3·52	1·18	24	5·59	3·23	31	3·40	1·42	62

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Persons	3	388·66	122·89
Conversations	8	16·17	2·02
Error	451	1694·62	3·67
Total	462		

F, persons/error = 33·49. *P* at 0·001 = 5·42 (for D.F. 3∞). ϵ = 0·985.
F, persons/conversations = 60·84. *P* at 0·001 = 15·83 (D.F. 388).

Comparing the capacity of respiration rate and ventilation percentage to discriminate between individuals both analyses of variance show the latter measure to be more sensitive in discriminating between persons (the *F* ratios for *V* % being 126·3 and for *RR* 29·6), and the second analysis of variance (Tables 3 and 4) also shows that the ventilation percentage discriminates more sharply (*F* = 43·3, *P* beyond 0·001) between the first and second part of the interview than the respiration rate (*F* = 2·62); the reliability coefficients ϵ calculated on the basis of variances were for the rate of respiration 0·957 and

Table 3. *Means and standard deviations of respiration rates of nine patients and three intelligent normal adults based on* 355 *utterances made in thirteen interviews*

Persons		First half of interview			Second half of interview		
		Mean	S.D.	*N*	Mean	S.D.	*N*
Patients							
P III (male)	interviewed by *B*	5·15	2·28	8	4·05	1·81	8
P VII (male)		9·12	3·59	11	10·48	3·32	11
P IX (male)		7·52	2·06	10	6·83	2·04	10
P X (male)		8·61	1·62	12	7·38	2·85	12
Ma (female)	interviewed by *A*	10·46	2·56	17	9·12	2·53	17
Nu (female)		9·78	2·18	24	9·46	1·72	24
Stn (female)		8·83	3·48	18	7·09	4·04	18
G (schizophrenic)		10·57	3·19	34	9·32	2·30	34
Normals							
S.O. (male)		3·88	1·37	6	7·10	3·62	5
S. I (female)		7·76	1·79	18	6·61	1·43	18
S. II (male)		9·49	3·81	20	8·33	2·39	20

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Between persons	10	839·59	83·96
Between first and second half of interview	11	82·09	7·46
Error	334	947·77	2·84
Total	355		

F, persons/error = 29·56. P at 0·001 = 2·74. $\epsilon = 0·983$.
F, between first and second half of interview/error = 2·62. P at 0·001 = 2·74.
F, persons/conversations = 11·25. P at 0·001 = 7·63.

Table 4. *Means and standard deviations of ventilation percentages of nine patients and three intelligent normal adults based on* 356 *utterances made in thirteen interviews*

Persons		First half of interview			Second half of interview			
		Mean	S.D.	*N*	Mean	S.D.	*N*	$dM_1 - M_2$
Patients								
P III (male)	interviewed by *B*	2·76	1·11	8	2·04	0·95	8	0·72
P VII (male)		3·56	1·07	11	4·12	1·40	11	2·16
P IX (male)		3·46	1·14	10	3·02	1·12	10	0·44
P X (male)		3·32	0·71	12	2·79	0·96	12	0·53
Ma (female)	interviewed by *A*	4·75	1·43	17	3·75	1·18	17	1·00
Nu (female)		4·93	1·13	24	4·81	1·30	24	0·12
Stn (female)		3·64	1·30	18	4·15	3·32	18	0·51
G (schizophrenic)		4·67	1·51	34	3·58	0·87	34	1·09
Normals								
S.O. (male)		2·08	0·70	6	2·80	1·16	6	0·72
S. I (female)		3·15	1·04	18	2·84	0·68	18	0·31
S. II (male)		5·13	1·93	20	4·35	0·72	20	0·78

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Between persons	10	921·88	92·19
Between first and second half of interview	11	347·92	31·63
Error	335	244·44	0·73
Total	356		

F, persons/error = 126·3. P at 0·001 = 2·74. $\epsilon = 0·996$.
F, between first and second half of interview/error = 43·33.
F, persons/conversations = 2·72 (P 0·05).

0·983 and for the ventilation percentage 0·987 and 0·996. This was in both cases due to the smaller error of the ventilation percentage, i.e. to the greater constancy of this measure; the comparison of the first and second part of the interviews shows that the trend of the change is clearly in the direction of a decreasing degree of ventilation as the interview moves on. This may be explained when we assume that normally interviewees settle down after a state of initial excitement or flurry to a more calm and better controlled state of mind. It may be different with resistant and rigid interviewees loosening up towards the end, as was the case with *S.O.*, or when a more stressful topic or topics of greater emotional import crop up only in the second part of the interview, as was the case with *Stn* when the subject of pregnancy and child-bearing touching upon the patient's anxieties was discussed.

Correlations between respiration and speech rate, as well as between speech rate and output of speech per expiration, were not calculated. The reason was that observation showed that within limits all kinds of combinations occurred—and this can be seen from Fig. 2—as was envisaged on the basis of the mathematically possible mutual and reciprocal variations of the two variables in speech rate. The latter tended sometimes to be accelerated more by increasing the output per expiration, and the respiration rate varied little; this state of affairs would yield a positive correlation of speech rate and output of speech per expiration, and an insignificant correlation of speech rate and respiration rate. In some other cases the more responsible factor for the acceleration of speech would be an increase in the rate of breathing, with speech output per expiration varying within more narrow limits.

These proportions of the two factors in speech rate varied not only in the same person, but from utterance to utterance within the same conversation. Clear tendencies in one or the other direction were not usual in our sample, although where they appear they would classify a conversation as a whole according to the processes involved in its production.

V. DISCUSSION

The varying proportions in which respiration rate and output of syllables per expiration combine in speech rate indicate that different states of the organism may underlie the same rate of speech production. We may therefore expect also different psychological states to go with the different ingredients of breath rate, and of speech output per breath (number of syllables per expiration).

(i) *Psychological significance of breath rate*

Experimental evidence, as reviewed by Clausen (1951) and Altschule (1953), supports the conclusion that while respiratory effects vary markedly in degree and character, hyperventilation is the common accompaniment of strong emotion and mental disturbance. Studies of the relation of respiration and thinking (Golla, 1921; Schilling, 1929; Suter, 1912) show a picture of reduced breathing activity, breath holding during moments of attention, delays, or irregular breathing. Psychoanalytic writers have frequently stressed the relation between anxiety and breathing (Fenichel, 1953; Freud, 1936; Harnik, 1930; Kubie, 1948). Thus Fenichel (1953) reports that: 'Analysts daily observe passing stoppages of breathing going along with an increase in castration anxiety.... The impression is given that, in inhibiting breathing at an intended new action, at a perception, at a change of attention, the ego tests out whether or not it needs to be afraid.'

The amount of respiratory activity during speech should therefore be expected to indicate strength of affect as far as it finds expression in speech. Easy and free-flowing affect would manifest itself in ample ventilation, excitement in hyperventilation, whereas repressed affect, attention, caution, and fear would be revealed in inhibited breathing.

The rate of respiration by telling us how frequently the tidal air has to be restored informs us about the amount of breathing activity. Although to make this information complete the further datum of the volume of tidal air is required, the rate of respiration has been found to be a relatively efficient measure of ventilation (Altschule, 1953; Ax, 1953), and more significant in discriminating between emotional states than an index of volume composed of rate times amplitude (Ax, 1953). Amplitude alone or the Inspiration/Expiration ratio were not significant at all.

(ii) *Psychological significance of the number of syllables per expiration*

The following facts from the physiology of speech production concern us here: (*a*) the energy necessary for the production of articulate sound is supplied by the expiratory phase of the breathing cycle; (*b*) the current of expired air can be modified at will, owing to the cerebral connexion with the respiratory centre; and (*c*) ordinarily expiration is a mere cessation of the active process of inspiration. While the latter cannot be much prolonged—it being normally checked by the Hering-Breuer reflex—expiration takes place spontaneously and requires only to be regulated to form a continuous current of air of sufficient duration for the production of sound.

It is possible to increase this outgoing current of air either by expelling a stronger current of air following a moderate intake of breath, by means of special expiratory activity involving muscular pressure, and so emptying the lungs, or by taking a deeper breath and so filling the lungs with a greater volume of air. An increased activity of the respiratory musculature of a voluntary and controlled kind is therefore necessary for both emphatic speech and acceleration.

Two aspects of the respiratory activity are thus relevant to speech production, namely (Meyer, 1892): (*a*) the rate of escape or expulsion of the outflowing current of air, or reciprocally the degree of ventilation of speech (i.e. the amount of air escaping through speech); and (*b*) the neuro-muscular activity controlling the action of the respiratory musculature and thus regulating the rate of escape of the expelled air current.

We are able to modify the strength and duration of the outflowing current of air, allowing it either to escape slowly and evenly as in normal speech, or at the other extreme to be emitted quickly and violently as in the case of a shout. Between these two extremes we must conceive of the rate of expulsion as a continuum covering all degrees of air escape from the lowest to the highest possible, ranging from the most economic use of the outflowing air current under strict muscular control to the quicker escape of air in a state of relaxation.

These physical processes provide the operational terms by means of which certain qualities of expression in speech, which may range from the monotonous and dry, through the warm and emotional, to the positively effusive and excited gush of speech, are becoming tangible and measurable. This assumption is supported by Stetson's (1928) findings resulting from his studies of speech movements in action. Stetson showed the divisions of the syllables to be identical with the divisions of the chest pulses one from another by demonstrating experimentally that the syllable movement equals the 'breath pulse' which equals the 'chest pulse'.

Thus the syllable has been shown to be not only a phonetic unit but also a unit of respiratory activity, i.e. a motor unit of the respiratory apparatus. *The number of syllables uttered during each expiration would therefore measure the rate of expulsion or escape of the outflowing air in speech.*

In terms of the process measured, the more syllables spoken during an outlet of breath the slower the pace at which the inspiratory muscles pass from a state of contraction to that of relaxation, i.e. at a constant speed of talking. What is measured is therefore the delay of the relaxing phase of the respiratory cycle, effected by voluntary control, a motor phenomenon which seems analogous to Luria's (1932) delayed action as controlled by his 'functional barrier'.

The actual length of the contraction-relaxation phase depends on the rate of respiration as well as on the rate of speech. For the number of syllables uttered per outlet of breath tells us how far the energy supplied by the outflowing air current was under cortical control, i.e. how far it has been guided into the channels of speech activity, and how efficiently it has been utilized, i.e. into how many speech units it was transformed. Or, by reversing our viewpoint, we learn from it to what extent the air current returned free and unbridled, as a spontaneous and unexercised carrier of emotions, as in laughter, sighs, shouts, etc., and what proportion of it escaped per speech unit (degree of ventilation in speech).

We may therefore expect that the number of syllables per expiration, i.e. the degree of ventilation in speech, will prove to be a measure of the extent of the transformation of the energy of expiration into verbal activity, that is of the extent to which the involuntary action of respiratory reflexes is directed by the cortical action of producing speech sounds.

The testing of these hypotheses by means of the above measures is in progress.

VI. SUMMARY

The aim of the investigation was to develop indicators of changes in emotional states based on the speech process itself. Previous research had suggested that respiratory behaviour in speech might be a fruitful line to study. A simple method of recording breath intake from sound recordings was discovered. Two measures were derived from speech-breathing activity: the rate of respiration during speech and the output of speech (number of syllables) per expiration. Both of them proved to be highly reliable and sensitive, but the output of speech per expiration was the more constant, as well as more sensitive, of the two. The phonetic and physiological facts of speech production and psychological evidence and theories, in respect of respiratory behaviour, suggested interpretations of the psychological significance of these measures which are in the process of being tested.

References

Altschule, M. D. (1953). *Bodily Physiology in Mental and Emotional Disorders.* Chapter: Respiration. New York: Grune and Stratton.

Ax, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosom. Med.* xv, no. 5, 433–42.

Clausen, J. (1951). *Acta psychiat., Kbh.*, Suppl. 68. Copenhagen: Ejnar Munksgaard.

Fenichel, O. (1953). *The Collected Papers. First Series.* New York: W. W. Norton.

Freud, S. (1936). *The Problem of Anxiety.* New York: Norton.

Goldman-Eisler, Frieda (1952). Individual differences between interviewers and their effect on interviewees' conversational behaviour. *J. Ment. Sci.* xcviii, 660–71.

Goldman-Eisler, Frieda (1954*a*). A study of individual differences and of interaction in the behaviour of some aspects of language in interviews. *J. Ment. Sci.* c, 177–97.

Goldman-Eisler, Frieda (1954*b*). On the variability of the speed of talking and on its relation to the length of utterances. *Brit. J. Psychol.* xlv, 94–107.

Golla, F. L. (1921). The objective study of neurosis. *Lancet*, ii, 115–22.

Harnik, J. (1930). One component of the fear of death in early infancy. *Int. J. Psycho-Anal.* xi, 485–99.

Kubie, L. S. (1948). Instinct and homeostasis. *Psychosom. Med.* x, 15–30.

Luria, A. R. (1932). *The Nature of Human Conflict.* New York: Liveright.

Meyer, G. H. (1892). *The Organs of Speech.* London: Kegan Paul Trench and Trübner.

Mowrer, O. H. (1953). *Psychotherapy, Theory and Research.* New York: Ronald.

Ruesch, J. & Prestwood, A. R. (1949). Anxiety, its initiation, communication, and interpersonal management. *Arch. Neurol. Psychiat., Chicago*, lxii, 527–50.

Schilling, R. (1929). Über 'Inneres sprechen'. *Z. Psychol.* iii, 204–55.

Stetson, R. H. (1928). Motor phonetics, a study of speech movement in action. *Arch. néerl. Phon. exp.* iii, 1–216.

Suter, J. (1912). Die Beziehung zwischen Aufmerksamkeit und Atmung. *Arch. ges. Psychol.* xxv, 78–150.

Travis, L. E. (1931). *Speech Pathology.* New York and London: Appleton.

(*Manuscript received* 20 *July* 1954)

CORRESPONDENCE

The Editor
British Journal of Psychology

Dear Sir,

I would like to comment on the paper by Prof. Oldfield.[1] If I appear to write *ex cathedra* and without explanation of terms, the limitations inherent in a letter must be blamed.

The analogy of an inscription on stone is not inappropriate to the initial registration of events. Moreover, it is quite possible for such an inscription to wear away in such a manner that, at various stages, it appears to be in different 'scripts' and thus to have different meanings. This would be an example of 'disconnected ambiguity'[2]—an important phenomenon related to the dimensionality of the coding. The real objection to the 'inscription on stone' type of analogy is the implicit, but not necessarily true, assumption that it does *not* give rise to this kind of ambiguity during erasure. But disconnected ambiguity is a striking feature of perceiving and remembering, and I suggest the time has come when we can make these more fundamental concepts explicit.

It is true that the known functions of regenerative loops in the central nervous system 'relate to the temporal prolongation of inhibitory and excitatory effects'. The point is that these prolongations represent storage of information, and hence a crude but genuine form of memory. Probably the physiological techniques applied to such loops are incapable of detecting any form less crude.

The 'circulating memory' employed in computers does not, in any relevent sense, embody a 'quite novel principle' (contrary to the author's assertion). The essential principle is that of preserving the message by repeatedly rewriting it before it has become unintelligible, just as ancient documents may be copied or archeological relics photographed before they crumble into dust—or as we deliberately rehearse a verbal message in order to fix it more firmly. It is worth remembering that the ordinary strategems and artifices of mankind will provide nearly all the qualitative models we need, without recourse to special technologies. The use of binary number codes is not a technology but an unexceptionable logical device, so far as we are concerned.

The statement that 'economy can be effected (by the type of coding proposed) without impairing accuracy of reproduction' is only true if there is no chance of error anyway—i.e. if the system is 'noise-free', in the jargon of information theory. Otherwise, minimum-redundancy coding (redundancy is a measure of the superfluity of symbols in a message) must be paid for by a greater frequency of errors. The nervous system is very far from noise-free, and we ignore the fact at our peril.

For example, the 'sudden' recoding which, as the author suggests, manifests itself as 'insight' need not be left, as he leaves it, to some mysterious and *ad hoc* 'active searching'. Given the necessary conditions for disconnected ambiguity, it will come about sooner or later with the same kind of probabilistic inevitability as ensures the increase of entropy in a closed physical system.

---

[1] Oldfield, R. C. (1954). Memory mechanisms and the theory of schemata. *Brit. J. Psychol.* XLV, 14–23.

[2] Woodward, P. M. (1953). *Probability and Information Theory, with Applications to Radar.* London: Pergamon Press.

As a further *ad hoc* hypothesis, the author invokes a special feedback mechanism to minimize redundancy. Possibly he did not intend it as *sui generis* and distinct from the basic mechanism of learning. Be that as it may, a simple way of looking at the matter is to note that, since the greater the redundancy the greater the protection against 'noise', that form of coding which will longest withstand the attrition of such random disturbances is the one which attributes the least information to the original message. Therefore, given the necessary coding filter for successively lower levels of information (increasing the redundancy attributed to the original message), the tendency to minimize the storage space occupied is automatic.

Incidentally, diminishing the storage space occupied is not to be understood as reducing the number of symbols in the store, but as reducing their total entropy. When you clear the register of a desk computer, it contains exactly as many symbols as before—only they are all noughts.

The formation of the encoding filter (or scheme of classification, as it may more intelligibly be termed) can, at this level of abstraction, be best regarded as the familiar operation of averaging—i.e. accumulating measures of statistical frequency, which act as 'probabilities'. A frequency-count destroys (by ignoring) an enormous amount of information; consequently it would appear that the local entropy is reduced at the cost of a vast increase of overall entropy, as general principles require.

The elementary processes of attrition and accumulation, together with disconnected ambiguity, seem, therefore, to be capable of exorcizing most of Prof. Oldfield's ghosts.

Yours faithfully,

W. E. Hick

Psychological Laboratory
University of Cambridge
3 March 1954

PUBLICATIONS RECENTLY RECEIVED

The Unconscious Origins of Berkeley's Philosophy. By J. O. Wisdom. London: Hogarth Press. 1953. Pp. xii + 244. 25*s.* net.

The kernel of this book is an attempt to apply to Berkeley's philosophy a psychoanalytic interpretation. Wisdom uses current psychoanalytic theory of the British School to give a psychoanalysis of Berkeley and thereby present the psycho-genesis and function of his metaphysic. The central theme which emerges is Berkeley's infantile attitude to faeces, and how this led him to the doctrine of 'esse percipi'—a doctrine which cleansed the world for him of the external poison of 'matter'. I have no room here to present any of the details of Wisdom's analysis. It will have to be sufficient for me to say that Wisdom seems to have made a careful and interesting application of psychoanalysis, and one which analysts themselves will no doubt consider to be a contribution to the subject.

Now I shall pass over the objection that it is really not worth while attempting the psychoanalysis of the dead, since, in the absence of the effects of interpretations upon the patient, the whole enterprise becomes too speculative. I shall also pass over the general objection from the hard-headed members of the British Psychological Society, namely that the whole of Wisdom's analysis of Berkeley is just sheer phantasy anyway and not worth bothering about. What I shall do is to use this review to touch on one problem of importance for psychologists that this book raises but does not develop or answer. This is the problem of the relation between 'the causes of' and 'the reasons for' belief, between the psychology and the logic of thinking. Wisdom's book raises this problem in the following way.

Wisdom says that 'the chief aim of his book is to throw light on philosophy' by psychoanalysing Berkeley, the man and his philosophy. In other words, he offers us the key sources (or causes) of Berkeley's philosophical position, and claims that thereby he is doing something of philosophical relevance. However, I can only discover one paragraph (on page 230) in which he does try to tell us how his book 'throws light on philosophy', and unfortunately I find this paragraph so confusing that it fails to enlighten me. I suspect that other British philosphers will find themselves here in the same difficulty as myself. I think it is a pity that Wisdom did not go on to develop the philosophical relevance of psychoanalytical interpretations, and thereby show the bearing that psychological 'causes' have on 'the reasons' in which philosophers are interested. For, in my opinion, the analysts have got hold of something here which is well worth developing. A brief review is hardly the place to exhibit the relevance of these psychoanalytical interpretations, but I shall nevertheless try to indicate one way in which this may be done.

Consider Berkeley's arguments for the view that 'esse percipi'. When we come to understand this view, we also appreciate that, far from being false or fantastic, it is defensible and true—to put the point crudely. Yet most of us at present have no wish to talk like this, and contemporary philosophers would maintain that we are not logically obliged to do so either. But when Berkeley argued that 'esse percipi', he quite obviously thought that by his arguments he *had* provided 'a good reason' why we are logically obliged to talk like this. To-day we can appreciate the force of Berkeley's arguments, but we do not accept them as constituting a good reason for having to say that 'esse percipi'. Why, now, did Berkeley accept his arguments as constituting 'a good reason'? In answering this question we have to ask why this way of talking—namely, 'esse percipi'—resolved his own philosophical conflict. If Wisdom is right, part of the answer will be that this way of talking helped to resolve his philosophical difficulty about matter, and it did so because this conflict was, in part, the product of a psychological disorder which had its origins in his childlike attitude to faeces. To talk of 'esse percipi' satisfied him partly because it helped to resolve the psychological problem from which he suffered and which contributed to form the character of his philosophical difficulty. His implicit criterion of 'a good reason' in this context is determined in part by his psychological disorder and his early infantile attitude to faeces. If Wisdom is right, he has drawn attention to certain unconscious sources of Berkeley's philosophical difficulty and of his compulsion to resolve it in the way that he did.

We can now point to the philosophical relevance of psychoanalytic interpretations. They are relevant to the theory of philosophical activity, since a description of what philosophers are trying to do has to allow for the fact that they are trying to resolve philosophical difficulties which may sometimes be generated in part by sources of the type with which psychoanalysis has familiarized us. This suggests that there may not be one 'right' or 'correct' way for everyone of answering a philosophical question. It suggests that the appraising terms of philosophical discourse, like 'good reason', 'right (or wrong) in saying this', 'mistake', 'true', 'false', 'correct', and so forth, need an elucidation that brings out their personality dependence. When, however, we appreciate the theoretical relevance of psychoanalytic interpretations, the manner in which we *practise* our philosophical activity is likely to be affected in various ways. Consider one such way. Suppose I point out—as a contemporary linguistic philosopher may do—that 'much of Berkeley's theory does not contradict common sense' (Warnock, *Berkeley*. Pelican Books, 1953). If I do *not* appreciate the theoretical relevance of psychoanalytic interpretations, I may easily imagine, as many contemporary philosophers appear to do, that this statement about Berkeley's theory is true in the same unqualified or universal way as is the statement that 'Everest is the world's highest mountain'. But if I come to appreciate the point of psychoanalytic interpretations, I am *less* likely to imagine that my statement about Berkeley's theory is true in the same straightforward, unqualified way as is my statement about Everest. For I can now appreciate that my statement about Berkeley, in contrast usually with the statement about Everest, may be part of an attempt to deal with a psychological conflict of my own, or with one fairly common at the present time in this country among the products of the Public Schools. It may be that what makes me want to emphasize this statement about the commonsensical character of Berkeley's theory and feel satisfied when I have shown its force is, in part, my own compulsive adherence to the ordinary speech learnt at my mother's knee; and this in turn may go back to an infantile wish to return to the ordinary milk and secure breast of my mother. It follows that for those who are *not* in a psychological conflict similar to mine, my statement about Berkeley may be uninteresting, pointless and not at all satisfying. To such people my statement about Berkeley will be true but not obligatory, since I have not given them 'a good reason' for talking in this way about Berkeley. Consequently, when I appreciate the point of psychoanalytic interpretations, I am likely to use appraising terms like 'good reason', 'mistake', and the like, much more cautiously and in a way which tries to make explicit my own criteria governing their use. This may have a further consequence. It is well known that much philosophical activity in England at present consists of a naturalistic exploration of our speech and language—a pursuit which is apt to degenerate into dictionary mongering and armchair social anthropology. When we appreciate the point of psychoanalytic interpretations, we may cease making this pursuit the central, or dominant, feature of our philosophical activity. We may be helped to recall that philosophical activity is only worth while if it is interesting, and that it is only interesting if it deals with philosophical *difficulties* which *really* trouble us.

But of course, to say all this is merely *to touch* on the philosophical relevance of psychoanalytical interpretations. I have not even pointed to their obvious limitations, namely that they do not deal with those unconscious sources of our philosophical difficulties which are to be found in our verbal habits—sources which the work of Wittgenstein and John Wisdom have emphasized, and which, unlike those psychoanalysts point to, have an immediate and critical relevance for philosophical activity. It is because this whole theme is exciting and obviously important that I am sorry Wisdom did not take the opportunity provided by this book to give us his views on the philosophical point of presenting us with the unconscious origins of Berkeley's philosophy. If he had done so, he would certainly have helped to clear up the current confusion in theoretical psychology over the relation between the logic and the psychology of thinking and belief.

B. A. FARRELL

Research Methods in the Behavioral Sciences. Edited by L. FESTINGER and D. KATZ. London: Staples Press Ltd. 1954. Pp. xi+660. 50*s*.

Books on research methods tend to be rather tedious, particularly in fields where there is little agreement on methodology. Books written by a large number of different authors tend to be repetitive, lacking in uniformity of style, and evincing no clear-cut plan or design. *Research Methods in the Behavioral Sciences* is singularly free from both these faults. It is well planned and put together, the individual papers complement each other without overlapping, the style of writing is uniformly good throughout, and the contributors almost without exception have written at a very high level of professional competence indeed. The book is likely to become the standard text in this field for many years to come.

One of the reasons for this unusual success derives, undoubtedly, from the fact that all those concerned in this production are associated with one university. The University of Michigan, through its Institute of Social Research, which combines the divisions of the Social Research Centre and the Research Centre for Group Dynamics, was in an unusually favourable position for undertaking a task such as this, due to the exceptionally large number of competent and experienced social psychologists, sociologists, and other social scientists working there. Much of the material reviewed in this book consists of original research, frequently unpublished, of members of the Institute; this intimate acquaintance with, and appreciation of, the complexities of social research is clearly reflected in the very sophisticated discussion of the problems which arise. It is difficult to single out chapters for special praise. To the reviewer, Helen Peak's 'Problems of Objective Observation' and Clyde Coombs's 'Theory and Methods of Social Measurement' appear particularly outstanding, the former giving the best discussion to date of different types of measuring and scaling devices, the latter putting forward a new, and in many ways, highly original theory of psychological measurement.

For the serious student, this book will certainly become required reading, and even the expert will learn a good deal from it. To British readers who may not always receive the duplicated or printed research reports so extensively quoted in this book, it may in addition serve as a reminder of the many different areas in which social psychologists in the United States are at present carrying out research. They may also feel that the almost complete absence of such research from the British scene does not reflect favourably on the prestige of British Psychology.

H. J. EYSENCK

Experimental Psychology (revised edition). By R. S. Woodworth and H. Schlosberg. New York: Holt and Co. 1954. London: Methuen and Co., Ltd. 1955. Pp. xii + 948. 63*s.*

It is sixteen years (1938) since the first edition of Woodworth's *Experimental Psychology* was published, and every psychologist who has used, read, or referred to this valuable survey of laboratory psychology will rejoice that a second edition is now available and that Prof. Woodworth himself has been able to play a part in its preparation. The new edition has about the same number of pages as the old, but the use of double columns has enabled about 20% more matter to be squeezed into it. A great deal of new material has been included. In fact, nearly half the references are to the work of the last 15 years and over 500 of them to books and articles published since 1950. Prof. Woodworth and his co-author, Prof. Harold Schlosberg of Brown University, have done a good job of revision.

Unfortunately some material from the older edition has had to be omitted. The chapters on feeling and on experimental aesthetics have gone altogether: accounts of older work have been curtailed. The amount of space, for example, devoted to the older work on memory has been considerably reduced. Hence, as the authors are at pains to point out in their introduction, the new edition does not altogether supersede the old, which still remains valuable for reference. The principal additions, apart from the fresh material added to every chapter, are an entirely new chapter on scaling methods and a greatly increased treatment of animal learning. The inclusion of the latter was probably inevitable if the book is to serve the needs of American classes, but it is not full enough to be really useful except to the undergraduate student, and it duplicates a great many other surveys of learning and learning theories. Since the authors have endeavoured to retain the 'style, viewpoint and general level of difficulty' of the first edition, a case could have been made out for excluding animal experiments, which were only very briefly mentioned in the earlier volume.

In general, however, Woodworth and Schlosberg have succeeded in their aim. This is still recognizably the same Woodworth in a rejuvenated form. Perhaps it seems on the whole rather more elementary in its treatment than the older edition, due no doubt to the greatly increased material which has been compressed into not very much greater space, and the necessarily rather cursory treatment of many topics, where progress in knowledge has been rapid. And possibly the eclectic approach which it adopts seems somewhat less satisfying to a generation that has become much more interested in and impressed by the importance of theory. Still the stuff of experimental psychology is here—here are accounts of a great many important recent investigations, by Michotte, Ames, Gibson, Sherif and dozens of other contemporary workers.

L. S. HEARNSHAW

L'Année psychologique. 53ème année, 1953. Paris: Presses Universitaires de France. 1500 fr.

British psychologists are inclined to feel a little indignant when their American colleagues cannot spare the time to sample the good things which, for all they know, abound in our publications. Yet our own attitude to Continental psychology is remarkably similar. *L'Année psychologique* is a convenient means of reducing our ignorance, at least as far as France, probably the most psychologically active of our near neighbours, is concerned. The latest issue follows the usual pattern: we have some original articles, some reprints of lectures delivered before the Association de Psychologie Scientifique de Langue Française and summaries of psychological articles and books that have appeared recently.

The summaries show not only that French psychologists are diligent in scanning British and American products but also that they are better placed than ourselves for observing developments in Central Europe. But the topics selected and the space allotted to them are strikingly different from what we should expect in comparable English language reviews and reminiscent of what they would have contained some decades ago. 'General psychology' is granted less than a quarter of the pages in question. There is much discussion of detailed analyses of sensory and motor processes but little of the central processes that link the two, apart from the contributions of neurophysiologists and psychoanalysts. The areas that most British and American general psychologists would consider of most moment—learning, motivation, perception, symbolic functions, theory-construction—receive little or no mention. There is little evidence of interest in the various efforts that have been made, in the U.S.A. and other English-speaking countries, to use systematic theoretical models to integrate the various fields of psychology and to guide research. Piaget's paper, in which he points out some parallels between his recent thinking and that of the cyberneticians and the Chicago mathematical biophysicists, is a thought-provoking exception. Reuchlin contributes a valuable article on research design, but he writes in a way suggesting that the topic is new and not accepted without suspicion in France.

Highly stimulating reviews of recent work in neurophysiology are provided by Piéron and Ajuriguerra. We can see from them what the French psychophysiological tradition must still command our respect. But they also make it clear to us once again that we cannot afford to wait for the neurophysiologist to solve all our problems for us and that the faster we progress in devising scientific psychological techniques for working out the laws of behaviour the better it will be for both of us.

This volume, in short, is one more reminder that psychology is still not an international enterprise like other sciences but that it is still all too neatly parcelled up into national allotments. But we can help to remedy this by adopting the habit of looking at *L'Anneé psychologique* and at the same time find much that is well worth watching in French psychology.

D. BERLYNE

The Design and Analysis of Experiment. By M. H. QUENOUILLE. London: Charles Griffin and Co. Ltd. 1953. Pp. xiii + 356. 36*s*. net.

In recent years at least half a dozen texts have appeared each presenting, at varied levels of sophistication, techniques suitable for the design and analysis of certain types of experiment. In view of their almost exclusive devotion to the methods of analysis of variance it is unfortunate that the titles of these works, like the one under review, tend to suggest that these methods are applicable to the analysis of data obtained from any type of experiment.

The first half of the text under review gives an account of methods suitable for analysing of data obtained from experiments designed in the form of randomized blocks, Latin squares, factorial schemes, and incomplete blocks. These are topics which may be found in most books on the subject but readers with no previous acquaintance of statistics will find difficulty in following the rough intuitive arguments that are given as justification for the methods of analysis. From this point of view the work suffers in comparison with the well-known work by Snedecor.

The second half of the book breaks new ground. In particular, the author discusses methods by which we may estimate the residual effects consequent upon applying a sequence of treatments to the same subject. For the experimental psychologist, what can be done in this direction still appears to be rather limited, but the suggested techniques will be of great interest to all who are persuaded that analysis of variance procedures are invalidated if repeated observations are made on the same subject. Other chapters deal with techniques for combining the results obtained from several experiments, the appropriate analysis to be employed when observations are missing, and the various scaling methods available for transforming data into a form suitable for the application of analysis of variance.

The expert will find no clear account of the fundamental probability models underlying the various types of analysis nor any details of the least squares approach to the subject. He will, on the other hand, appreciate the discussions devoted to the more complex experimental designs which have been recently developed and which, up to now, could only be found in journals. Further, he will find that the examples given to illustrate the methods, though remote from psychology and drawn mainly from the agricultural and medical sciences, are discussed with commendable common sense.

In general the book appears to be suitable neither for the beginner nor the expert, but is invaluable for the treatment it gives of many specialized topics related to analysis of variance which are not to be found in existing texts.

A. R. JONCKHEERE

Traité de psychologie animale. By F. J. J. Buytendijk. Paris: Presses Universitaires de France. 1952. Pp. xv + 362. 1000 frs.

The title of this work is misleading; it cannot by any reasonable criterion be considered as a 'Treatise' on animal psychology. The author himself makes this clear in the introduction, remarking that he has omitted any treatment of tropisms, the behaviour of unicellular organisms, perception, and learning phenomena. To most readers it will seem as if some of the main topics in animal psychology have been ignored. But Prof. Buytendijk has other aims in mind than merely giving a systematic account of the subject. By means of discussions of topics such as laughter and play in animals and whether chimpanzees can 'lie', the author attempts to introduce a new approach to the study of animal behaviour. This new approach is unfortunately a type of phenomenological analysis popular in literary and philosophical circles on the Continent and which has now, with the work under review, reached the frontiers of animal psychology. The result is that a distinguished authority in the field has produced a book which by the opacity of its style (though it should be remembered that we have here a translation) and the looseness of its thought, will dismay readers who are acquainted with the author's earlier empirical work. The book may be recommended, however, as an example of one of the prevailing trends in contemporary Continental psychology.

A. R. JONCKHEERE

Physiologische Psychologie der Sprachvorgänge. By Friedrich Kainz. Stuttgart: Ferdinand Enke Verlag. 1954. Pp. 571. Paper, DM. 43.50; linen, DM. 47.00.

Prof. Kainz's book on the physiological psychology of speech continues a traditional line of Viennese psychology. It is, after Bühler's *Sprachtheorie*, the most important work in a field which has been much neglected by psychology at large. Between the recent revival of interest in the study of language behaviour and Wundt's emphasis on the utmost psychological importance of the phenomena of language lies a fallow period during which the study of language was neglected by psychologists and left entirely to those interested in linguistics, phonetics, and the physiology and neurophysiology of speech. An enormous body of basic knowledge has grown out of the activities in these fields, of which the general psychologist is largely unaware. With the growing interest in the subject of communication language behaviour demands more and more attention as an area of investigation.

Prof. Kainz's book is, at this juncture, invaluable to any student or worker in the field of linguistic psychology. It is marked by unusual breadth, is encyclopaedic in scope and immensely serviceable in as far as it puts before the reader a wealth of well-digested material. If, under these circumstances, the book has grown to awe-inspiring proportions, the advantage of the unstinted supply of information given is that it allows the reader to form his own opinion in cases of controversy, even when Prof. Kainz makes his viewpoint clear. No advantage, however, seems to accrue from the long-winded and long-worded style, even by German standards, which one fears will deter many a potential foreign reader with a working knowledge of German.

The volume reviewed here is the third of the four volumes comprising the whole work. In the two preceding volumes Prof. Kainz dealt with the properties and achievements as well as with the origin and development of language. In the volume under review he surveys all those peripheral and central processes which come into operation in the act of speaking or of understanding spoken language. Prof. Kainz sees his task to include the presenting of a comprehensive psychology of speech based on the physiology of speech processes, as well as to give an historical survey of the science and the problems of linguistic psychology.

The book is divided into five chapters. In the introduction the author defines the science of linguistic psychology as dealing with language as an emotional, mental and physical activity of man. Its subject is the totality of the psychophysical processes and functions which operate in the production of speech. Prof. Kainz distinguishes two processes: (i) the central or cortical activity of diction, and (ii) the peripheral process of articulation. Speech-physiology therefore deals with the processes in the cortex and the brain stem, the intra-cerebral paths participating in the speech activity, and the afferent and efferent pathways in the nervous system activating the muscles of the organs of respiration, phonation, and articulation.

The psychological aspect of speech lies, according to Prof. Kainz, in the conscious intention, the verbal imagery (acoustic and kinesthetic), the schemata of sentences and beginnings of phrases, which are completed according to the initial plan, and finally the control and guidance of articulation.

The second chapter deals with the problem of localization and attempts to answer the question of what are the processes activated in the cerebral cortex during speech activity. The author's position in this controversy is a compromise. Prof. Kainz prefers to remain within the principle of localization critically modified in the light of Gestalt view as represented particularly by Goldstein.

Chapter three deals with the activity of speech. We are taken through the detailed phases of the speech process, from the initial voluntary impulse to speak to the detailed movements of articulatory execution which follow automatically. Prof. Kainz makes a clear distinction between the phases of thinking, diction or 'inner speech', and utterance. Diction is not the behaviourist 'sub-vocal talking' equivalent to thinking, but the totality of processes and dispositions which is responsible for the formation of verbal symbols, and their grammatical and syntactical structuring.

This precision of analysis in pursuing the actualities behind each term is characteristic of the whole book. Such ground-work enables the experimental worker to formulate concrete and testable hypotheses.

Of particular interest to the experimental worker in the field of speech behaviour and communication are those parts of this chapter which deal with the act of speaking. The discussion includes the action of the peripheral organs of speech which take part in the three processes basic to speech production: the respiratory organs, the organs of voice production, and those of articulation. Prof. Kainz then discusses the co-ordination of speech and hearing, the organic basis of articulation, and the relation between psychology and phonetics. In this chapter alone more than 350 references are cited. It is a valuable experience for any psychologist whose work takes him into the fields of interviewing or psychotherapy, for example, to be made to realize how many aspects there are to the basic activity of speaking alone and how much there is to be known about it.

While all these need to be considered in any comprehensive science of communication it is well to take heed of Prof. Kainz's warning that the study of these phenomena can only become fruitful from the point of view of psychology if they are brought into relation with the consciousness of the individual speaker. Such emphasis is the more salutary when seen in juxtaposition to the fascination exercised on psychologists by the achievements of communication engineers and information theory.

A further chapter is devoted to the subject of the understanding ('verstehen') of speech. In the last chapter Prof. Kainz deals with the subject of conversation. Consistent with his general outlook which has emerged as one of dynamic functionalism moderated by a strictly biological orientation, he rejects the behaviourist theories of conversational speech as being a mere mode of social action. He insists that expression cannot be limited to the monologue but that it is one of the functions of conversation. Through conversation, Prof. Kainz impresses upon us, we can study linguistic activity in the fullest realization of its potential powers, as a concrete linguistic event containing and integrating all functions of speech, active and receptive.

Psychiatrists and interviewers who are using conversation as the vehicle and instrument for their work will welcome this conclusion, for they will feel that it carries conviction in terms of their own experience. At the same time Prof. Kainz has provided a solid background to the exact study of the linguistic process by showing into how many component processes the process of speech can be broken down.

While these facts might have been presented less turgidly, and the index system arranged better to suit the reader's convenience, these are minor points if measured against the merit of the book. Prof. Kainz's service to the scientific study of psychological phenomena based on the use of speech and of linguistic psychology cannot be emphasized enough.

F. GOLDMAN-EISLER

Communication and Persuasion. By CARL I. HOVLAND, IRVING L. JANIS and HAROLD H. KELLEY. Yale University Press; London: Geoffrey Cumberlege. 1953. Pp. xii + 315. 36*s*. net.

Like all good things, the present trend towards greater scientific rigour in psychology can be carried too far. There is a growing tendency to feel that even the most incontestable statement about human behaviour and motives must be supported by statistical evidence, and a good deal of time and money is being spent on the experimental demonstration of the obvious.

A few such demonstrations are included in the book under review, which reports the results of a co-ordinated research programme conducted at Yale University on 'the modification of attitudes and beliefs through communication'. Thus one experiment demonstrates that if we greatly value our membership of a particular group we are likely to be resistant to views which conflict with the group's standards; several others, that we are more likely to agree with a statement, and to judge it as 'fair' and 'honest', if it comes (or is believed to come) from a source we regard as reliable than if it comes from a source we suspect.

But it would be unfair to judge the book by these examples; not all the results reported are so predictable *a priori*, and some are of considerable practical interest. For example: a fear-arousing, gruesomely illustrated lecture on dental hygiene produced significantly more emotional tension than a less fear-arousing lecture, but had significantly less effect on the subjects' future behaviour; one-sided advocacy of a particular view produced less resistance to counter-propaganda than did a statement which took account of the arguments for the other side; students developed a more favourable attitude towards a point of view after they had carried out the assignment of presenting it as favourably as possible to an audience; the most 'persuadable' students tended to be shy, unaggressive and liable to depression, while the least 'persuadable' showed tendencies towards neurotic anxiety and obsessional symptoms.

Besides describing the experiments carried out by the research team, the book contains valuable summaries of other relevant work. Though the style is as graceless as the current American convention demands (e.g. 'Fear of social disapproval might give rise to strong facilitating motivations with respect to acceptance of persuasive communications') the book is by no means unreadable; and it is admirably produced and printed.

M. KNIGHT

The Study of Language. By JOHN B. CARROLL. Harvard University Press; London: Geoffrey Cumberlege. 1953. Pp. xi+289. 38*s*. net.

This is a specialized book and will appeal mainly to those with a professional interest in linguistics, but many psychologists will be interested in the chapter on linguistics and psychology. The author starts from the axiom that thought differs 'only in the matter of degree of complexity' (i.e. only in complexity?) from simpler response-processes such as pressing a key in response to a signal; and he makes a provisional attempt to fit 'language behaviour' into the framework of S–R and reinforcement-theory. To the present reviewer it does not seem a promising approach; but some of the younger generation may feel stimulated and enlightened to learn that when they talk they are making socially reinforced verbal responses, and that when they use verbally expressed hypotheses as an aid to problem-solving 'the selection of linguistically-coded hypotheses is a function of two factors: (*a*) the stock of verbal responses which the subject can call up to solve the problem, and (*b*) the relative strengths of these responses in relation to the stimulus values of the material presented'.

M. KNIGHT

Developments in the Rorschach Technique. Vol. I. *Technique and Theory.* By BRUNO KLOPFER, MARY D. AINSWORTH, WALTER G. KLOPFER and ROBERT R. HOLT. London: G. G. Harrap and Co. Ltd. 1954. Pp. xi+726. 42*s*.

A full appraisal of the eagerly awaited 'new Klopfer' must wait until both volumes of the work are available. Meantime, the percipient reader will note that the present volume, by itself, is nearly twice the size of the familiar 'Klopfer and Kelley' of 1942. This comparison not only indicates the immense amount of Rorschach research that has been carried out in these twelve years; it also reflects the increased care and clarity with which most of the basic concepts are set out, and the much fuller theoretical discussion now provided (Part Three, entitled 'Theory', extends to nearly 200 pages). Of probably greater value still is the series of explicitly stated 'interpretative hypotheses' introduced at the outset of the discussion of each scoring category or relationship among factors. The effect, for the beginner, is to make the interpretative significance of the various types of response both clearer and more convincing; for the experienced Rorschach worker it affords an opportunity of re-examining the theoretical bases of the technique, which he might otherwise have been liable to take for granted.

Certain changes in practice, foreshadowed in the 1946 *Supplement*, are now fully developed. Chief among these is the discarding of the $F+$ category in favour of 'Form-level rating', a device which now assumes major importance, but which to the present reviewer has always appeared clumsy and irrational, especially when the ratings, which are made on a scale which is discontinuous, are averaged. The whole question of quantification in Rorschach is one which still requires to be thrashed out; thus, percentages and 'normal limits' are frequently quoted, but almost inevitably with the proviso that below a certain minimum (e.g. of total length of record) these findings cannot be applied, a safeguard which cannot but cast doubt on the validity of the relevant interpretation *within* the permitted limits.

Other innovations in method include the introduction of 'Sequence analysis', and of an 'Analogy period' (between the Inquiry and the Testing-the-Limits procedure). The former is almost self-explanatory in its intention, and strikes one as a valuable tool but one difficult to apply without undue expenditure of time. The conduct and function of the latter are, unfortunately, poorly defined. Additional contents include a very fully reported case-study (nearly 80 pages), and a description of a 'Prognostic Rating Scale' designed to predict a patient's response to psychotherapy.

It is perhaps a pity that the new book could not have been so written as entirely to supersede the earlier work. Much of the exposition (particularly in the first chapter of all, on 'Administration') is couched in language (not to say jargon) which presupposes familiarity with Rorschach concepts, and certain examples and other material previously cited are not repeated. That being so, one is tempted to look for redundancies, but in practically every case one cares to test the new is an improvement on the old. Consequently, the present work becomes one's vade-mecum, with Klopfer and Kelley retaining its place on the shelf, for reference.

BORIS SEMEONOFF

The Graphomotor Projection Technique. By S. A. KUTASH and R. H. GEHL. Springfield, Ill., U.S.A.: Charles C. Thomas; Oxford, England: Blackwell Scientific Publications Ltd. 1954. Pp. xi+133. 27*s.*

In this test the subject is blindfolded and asked (in effect) to scribble freely, without trying to make anything in particular, for five minutes. This is repeated, after which there is an inquiry asking for recall of what has been done on the first attempt, the outlining of objects, etc., seen in the production and finally associations to each of the things seen or imagined. The inquiry is repeated for the second production.

The structural and other qualities of the production are classified according to carefully worked out criteria, and these features are related to the behaviour and inquiry data to build up a diagnostic and descriptive personality picture. A research is reported in some detail showing a large number of significant differences in the performances of normal and schizophrenic subjects and examples of normal and schizophrenic records and interpretations are provided.

Some claim is made for this projective technique that it is an attempt to realize an aim, which arose out of the Allport and Vernon Studies in Expressive Movement, that it is important to study movements not in isolation nor as single traits but rather in respect to the basic motivation of the individual. Although the authors for the most part adopt a holistic and dynamic approach, this claim is not substantiated. In the reviewer's opinion a projective technique designed for clinical use should be capable of throwing light on the nature of personality dynamics, conflicts and their causation in terms of concepts directly relatable to therapeutic work. The rationale of the test and the results demonstrated are in far too general terms for this purpose. As an aside one would also express doubts as to whether the test, in the way it is given, represents a task which patients can readily and meaningfully accept.

The book makes some thoughtful contributions to an interesting area of research; it may well be that it would have been more advisable at this stage to present the work as such, rather than as a clinical technique.

As with all books from this publishing company the production is serviceable and pleasing.

H. PHILLIPSON

Dynamic Psychiatry. Edited by FRANZ ALEXANDER, M.D., and HELEN ROSS. Chicago: The University of Chicago Press; London: Cambridge University Press. 1952. Pp. xii+578. 75*s.* net.

This volume offers 'to students of psychiatry a comprehensive view of dynamic psychiatry'. The editors believe that the dynamic trend, essentially 'the advancement of the study of psychiatry from a descriptive into an explanatory phase', is the result of the impact of psychoanalysis upon the whole of psychiatry. This book, however, is not an exposition of psychoanalysis.

The book is in three parts: Part I, Concepts of Dynamic Psychiatry; Part II, Clinical Psychiatry; and Part III, Influence of Psychoanalysis on Allied Fields.

Part I consists of four chapters: I, Development of the Fundamental Concepts of Psychoanalysis, by Franz Alexander; II, Dreams and Rational Behaviour, by Thomas M. French; III, History of Metapsychological Concepts, by Edoardo Weiss; IV, Personality Development, by Therese Benedek.

Chapter II is disappointing. It is adapted from the author's *The Integration of Behaviour*, but is too brief to stand as the major section on dreams in such a volume as this. Chapters I and III should be read together. Weiss writes briefly on the early history of the concept of the unconscious and traces changes in metapsychological concepts. The writing is a little overladen with psychoanalytic terminology. There is not the contrast in these chapters which one might expect from the titles. Since both authors write historically or developmentally, a joint chapter might have made better exposition.

To the psychologist the title of Benedek's chapter is somewhat misleading. Perhaps *Psychosexual development of personality* would give a better guide to its content. Benedek applies the concepts in the chapters by Alexander and by Weiss to the developmental phases from infancy to senescence. For Benedek 'the development of the personality is the unfolding of an innate *anlage*—constitution under the influence of the environment'. This is an excellent chapter, but the argument would be easier to follow with sub-headings.

Part II: V, Neuroses, Behaviour Disorders and Perversions, by F. Alexander and L. B. Shapiro; VI, Acute Neurotic Reactions, by Leon Saul and J. W. Lyons; VII, Emotional Disorders of Childhood, by Margaret W. Gerard; VIII, Contributions of Psychoanalysis to the Study of Organic Cerebral Disorders, by H. W. Brosin; IX, Psychodynamic Approach to the Study of Psychoses, by J. C. Whitehorn; X, Contributions of Psychoanalysis to the Study of the Psychoses, by H. W. Brosin; XI, Principles of Psychiatric Treatment, by M. Levine. Chapter VI is mainly concerned with the so-called 'war neuroses'. The literature is briefly reviewed: Freud, Ferenczi, Abraham, Jones, Simmel, Kardiner, Radó, Grinker and Spiegel; incidence, etiology, types of stress, symptoms, dynamics of symptom formation, prognosis and treatment are among the topics dealt with.

Chapter V is perhaps the most disappointing in the book. One would expect to find here the full richness of the dynamic viewpoint. Instead there are 22 pages to cover definition of neurosis, hysterical conditions, obsessive compulsive states, depressions, hypochondriasis, behaviour disorders, alcoholism, drug addiction and sexual perversions. There are only eight references, and Fenichel is not included. Chapter VII should, perhaps, be read after Benedek's chapter. Under etiology there is a welcome statement of complex causation, 'the way in which a child reacts to any trauma is determined by the total constellation of constitutional trend, his status of physical maturation, his integrative capacity, his previous experiences, and the type of trauma'. The problems are arranged according to developmental levels. Treatment, diagnosis and prevention receive brief mention. This is a well written and concise account. There is no reference to recent work on autistic children, nor to the emotional problems arising on a secondary basis, e.g. as an accompaniment to orthopaedic handicaps, rheumatic fever, tuberculosis, etc. This is also a deficit in the book as a whole.

Brosin's chapters VIII and X are scholarly surveys of the literature and are not limited only to psychoanalytic references. In chapter X he hopes 'for a much more intensive research-therapeutic programme which will justify current assumptions that a large causative component of the psychotic process is psychological'. It is a brief chapter but covers 139 references from the Burghölzli group (1902–14) to B. Lewin (1950) on elation, and Eissler on schizophrenia (1951).

Whitehorn's chapter IX and Brosin's X should be read as a unit. Whitehorn writes, 'a psychosis is a form of reaction, not a mere deficit or impairment'. This chapter is weakened by the argument from psychotherapy. Levine's chapter XI needs evaluation by the practising psychiatrist to whom it is addressed, but to the reviewer it appears excellent. Broad in outlook, blending the goal of the scientific with that primary goal of being helpful to the patient, it discusses treatment at various levels against a frame of reference based on six considerations: clinical diagnosis; dynamic diagnosis, genetic diagnosis, transference, counter-transference and treatment possibilities. He briefly classifies techniques of psychiatric treatment and details the various types of psycho-therapeutic procedures. He relates this to the therapist's training and skill; 'only an accredited psychoanalyst should do psychoanalysis'. Clinical psychologists have therapeutic functions, 'if his training is adequate [he] may do the same varieties of psychotherapy [as the P.S.W., i.e., with close supervision by psychiatrists, using some of the aspects of the techniques of supportive relationship and expressive therapy limited to conscious material], while working as part of a psychiatric clinic team, although his primary responsibility is in the field of teaching, research and the use of special diagnostic tests.' This chapter is especially recommended to the D.P.M. student and the clinical psychologist.

Part III: XII, The Psychosomatic Approach in Medicine, by F. Alexander and T. S. Szasz; XIII, Some Relationships between Social Anthropology and Psychiatry, by Margaret Mead; XIV, Clinical Psychology, by D. Shakow; XV, Animal Psychology and its Relation to Psychiatry, by D. M. Levy; XVI, A Review of the Influence of Psychoanalysis on Current Thought, by H. W. Brosin. The chapter by Alexander and

Szasz might profitably be read following chapter x. Wisely the authors avoid the reified term and write of the psychosomatic approach. They briefly review changes in concepts and discuss differences in present views, including that of the Chicago Institute. Application follows of these concepts to the problem of gastro-intestinal disorders, bronchial asthma, etc.

Chapter xiii by Margaret Mead contains no surprises for those who know her work; it is a thought-provoking survey along with speculation on her own research and that of others. Her final sections are on the problem of normality-abnormality and comment critically on Kardiner and Roheim. Though there is a reference to the work of Rivers with the Todas, and to Hose and McDougall, there is no reference to the 1924 papers of C. G. Seligman and Ernest Jones.

Shakow's chapter xiv is somewhat limited, perhaps because he defines clinical psychology as concerned 'with the psychological adjustment problems of the individual'. One might expect reference to experimental procedures, e.g. Ebbinghaus's saving method in relation to psychiatric problems such as Korsakow syndrome. He also follows the usual American error of dating clinical psychology with references only to Witmer and Healy. There are, however, valuable discussions of what questions (in the diagnostic realm) should be put to the psychologist to which he may legitimately be expected to have a reply. Testing is discussed in terms of controls set up to reduce the amount of dependence placed on the observer. Shakow points out how the unified psychological portrait of the individual patient depends upon sampling of psychological functions. His plea is for scientific sophistication in this sampling.

Levy's chapter xv argues the thesis that, having accepted 'our kinship with the animal world structurally and biochemically', we should not 'remain isolationist psychologically'.

Brosin's final chapter, xvi, supported by 150 references, refers to the numerous precursors of Freud, the growth of psychoanalytic influence on psychiatry in terms of persons, books and organizations, followed by summary references to its influences in psychiatry, sociology, anthropology, economics, religion and literature.

Asked to choose the ten best books in clinical psychology and psychiatry published in the last 10 years, the reviewer would have little hesitation in including this volume. The D.P.M. student will find it valuable and not likely to go unused, if only as a bibliographic source, when he is a consultant. The clinical psychologist should be familiar with much of the theoretical chapters but should profit considerably from part II and the chapter by Alexander and Szasz. The index contains no reference to Heredity. There are several saving references scattered in the chapters to hereditary constitution, to universal inheritance, and some mention of individual variation, but genetics on the whole is ignored. There is no mention of the evidence for inheritance in manic depressive psychosis and schizophrenia. The psychologist is likely to feel that the volume would be better for drawing on developmental psychology and the psychology of individual differences, and for more adequate use of the concepts of dispersion and the significance of differences. There is no discussion and little reference to criticism of the dynamic hypotheses. Some personal expectations of the volume were not fulfilled; there is nothing on suicide, experimental psychosis or ecstasy states.

J. C. KENNA

Group Dynamics: Research and Theory. By DORWIN CARTWRIGHT and ALVIN ZANDER. London: Tavistock Publications. 1954. Pp. xiii + 642. 35*s*.

This is an important, comprehensive, and in places difficult, book. There are thirty-five well-chosen papers, and fifty-one contributors. It represents a well-integrated and frontal attack on the problem of application of scientific method to the study of group behaviour. While much of the book is informative and relatively easy reading, in places it is quite obviously the work of theorists, writing for other theorists, in a sophisticated and unavoidably technical way. As a contribution to psychological science the book has been greatly strengthened by the editors' decision to exclude papers which had 'no clear relation to any explicit theoretical conceptualization'. The research papers are classified in six sections, each section being introduced by a chapter written by the editors. The sections comprise: (1) Theoretical; (2) Group Cohesiveness; (3) Group Pressures and Standards; (4) Group Goals and Locomotion; (5) Group Structure; and (6) Leadership. A reasonably wide range of theoretical orientations is represented. Thus the first section includes papers by five theorists representative of somewhat different approaches. They include Cattell writing of group syntality from a factorial standpoint; Jennings as representative of Moreno's Sociometry; and Scheidlinger whose approach is psychoanalytic. Nevertheless, one would anticipate, in a book of this kind, considerable influence of the theory, concepts and researches of Kurt Lewin. This is broadly true, though only one paper by Lewin himself is included: a report on his classical studies of the effect of group decision on the changing of food habits. The editorial chapters are written from a standpoint which might be described as a *development* of that of Lewin. The italicized word is important, since the book as a whole is a record of the considerable development that has occurred in recent years, not merely of research, but also on the methodology side. If the editorial standpoint is 'Lewinian', it derives from Lewin in his later life: the *Field Theory in Social Science* period, not of the early 'topology' period. An opportunity appears to have been missed for taking account of major advances in recent years, in the understanding of the learning process. It is, for example, difficult to understand how the editors, in chapter 7, can discuss the increase and decrease of attractiveness (valence), of a group for its members, without reference either to the principles which govern reinforcement, or the process of extinction. Again, Lewin himself in his later years achieved close integration with the thought of the Neo-Gestalt wing of students of learning. In view of this it is disappointing that even Tolman finds no mention in the 642 pages of the book, other than (chapter 16), as a place name within the Massachusetts Institute of Technology.

As more detailed illustration of the content of this book we may instance the well-written editorial chapters on the six areas of 'group dynamics' examined; the interesting account by Schachter of the neglected problem of committee behaviour; several important chapters by Festinger and others on communication; investigations reported by Horwitz on the Zeigarnik effect in the field of remembering; and the intensive study by Gordon of a small group in terms of a single attitude, and its modes of expression. It is fortunate that a British edition of this book is now available and, in terms of its length, it is excellent value. There is much that can be read with profit by the student or intelligent layman, who can skip the difficult parts which the professional psychologist will have to 'work through' rather than merely 'read'. The volume as a whole contains many interesting and significant empirical observations. It represents also an account of theoretical and methodological advances which British Psychology as a whole cannot afford to ignore.

PETER MCKELLAR

Proceedings of the Association for Research in Nervous and Mental Disease. (December 14 and 15, 1951.) London: Baillière, Tindall and Cox. 1953. Pp. xiv + 451. 68*s.* 6*d.* net.

This latest symposium, issued by the Association for Research in Nervous and Mental Disease, is now published in Great Britain.

The work is arranged in seven sections. Each section comprises four or five papers by authoritative speakers, followed by a discussion, and a presidential criticism and summary of the several contributions to the sectional theme.

Four sections are devoted to individual and group psychotherapy; two sections concern biological, electrical and pharmacological treatments, and two sections are given to neurosurgery.

The sections on psychotherapy combine recent advances with an historical account of almost every psychotherapeutic method and its field of application. This material is presented in a very readable and digestible form. No one technique is given prominence over others. Each technique, be it long-term or short-term treatment, individual or group treatment, hypnosis, psychosomatics, environmental manipulation, etc., is described with emphasis on its range of application and its limitations. Here one finds no mere account of the birth pangs of psychotherapy. The subject is presented as a mature science in which formerly conflicting theories and warring 'schools' have now developed into essential pieces of the total pattern. Thus, the student who is in doubt as to the range of application of various psychotherapeutic methods will find this book a really practical help.

The sections which deal with insulin treatment, electrical treatment and various pharmacological treatments are the least impressive part of the book. Of course there is little to say about insulin and E.C.T. over and above what one finds in almost any text-book nowadays, but even the psychological effects of cortisone are discussed in a turgid style, and one gets the impression that this part of the work is a piece of not very seasoned stuffing which has been thrown in for luck.

The final two sections on neurosurgery are very good. Here one finds in a very condensed form all the latest information on the effects of the various cutting operations for the relief of psychological symptoms.

OTHO FITZGERALD

Psychological Problems in Mental Deficiency. By SEYMOUR B. SARASON. New York: Harper. Pp. x + 402. 40*s.*

This timely study of the difficulties of diagnosis of mental deficiency and its social and educational concomitants is as appropriate to the British scene as it is to the American. Work with mental defectives, although exacting, has often seemed unexciting and limited to ensuring that no physical or social disaster occurs; this book shows that, for the defective, classification *per se* does not answer the real questions, that the limits of realization and development that can be achieved have not yet been reached and that psychotherapy may well be a profitable method of treatment.

There is included an analysis of the varieties of amentia and particular attention paid to the 'garden' subcultural variety—which approximates to simple feeblemindedness—for the environmental causation of which Dr Sarason makes out a good case, setting it in the wider framework of the relations between child-rearing practices and behavioural patterns, and illustrating his point by examples of the considerable facility by which individuals move in and out of classification as feebleminded.

Dr Sarason makes a careful and well-documented study of assessment procedures and test interpretation; he will know that he takes most people with him when he qualifies the value of a particular figure obtained on an intelligence test, but he also indicates the danger of using other tests such as the Rorschach, for this purpose, on the presumption that they yield all the relevant data for assessm nt purposes—the best test may reveal little of the potential of a subject or how he functions in other situations.

Dr Sarason sees the problem of the defective not merely as a social or educational nuisance but as one of adjustment that the individual concerned, his family and his social milieu have to make, and not the least important chapter is the one in which communication of information to parents is analysed and sample conversations set forth. Many misunderstandings, gaucheries and insensitive pronouncements will be avoided in the future if the problem is approached on Dr Sarason's realistic basis: 'The handling of the problems requires not only knowledge of certain facts but an understanding of what is essentially a psychotherapeutically-oriented study; a sustained attempt on the part of some trained person to understand the motives of parents, their frustration and hopes, and by virtue of such understanding as well as by previous training to enable parents to accept more realistic and satisfying attitudes.'

The least successful chapter of the book is that concerned with brain injury and aphasic conditions; for example, no reference is made to the work of Head, Adrian and Zangwill, but this is the only largely theoretical excursion in what is an essentially practical book. It can be confidently recommended to all clinical and educational psychologists and to psychiatrists, medical officers and social workers who are concerned with mental defectives and their parents.

JOHN M. GIBBS

Isaac and Oedipus. By ERICH WELLISCH. London: Routledge and Kegan Paul Ltd. 1954. Pp. xi+131. 15*s*. net.

This short study of the father–son relationship represents a re-evaluation of the Oedipus Complex in terms of the Biblical story of the Akedah (the unfulfilled Sacrifice of Isaac by his father Abraham). Wellisch's approach differs from Freud's in several respects. First, whereas Freud formulated the Oedipus Complex from the son's standpoint, Wellisch considers it in terms of the mutual relationship of father and son. Secondly, the primacy accorded by Freud to patricidal motives in the son is accorded by Wellisch to infanticidal motives in the father. Thirdly, whereas Freud's formulation was based on Greek mythology, Wellisch draws upon a Biblical story registering the historical abandonment of the ancient practice of infanticide in favour of a sacrificial rite in which the son can participate with the father. Fourthly, whereas Freud describes the resolution of the Oedipus conflict in terms of an identification of the ego with an introjected parental figure in the form of the super-ego, Wellisch regards this resolution as only partial and precarious unless there is included in the super-ego the introjection of an objective moral force promoting altruistic love ('the voice of God' which arrested the sacrifice of Isaac). Thus, in terms of the Akedah story, Wellisch envisages the possibility of 'salvation' from the Oedipus conflict in contrast to the stoical 'compromise solution' envisaged by Freud in conformity with the fatalism of the Greek tragedy of Oedipus. Whatever view the psychological reader may adopt towards the religious aspect of the thesis expounded in this book, he can hardly fail to be impressed by Wellisch's exposition of the psychological importance of infanticidal motives, supported as this is not only by rich anthropological evidence, but also by the internal evidence of the Oedipus myth itself (for Oedipus was exposed upon a mountain at the instance of his father's jealousy long before he killed his father). According to Freud patricidal, combined with incestuous, motives are the inevitable outcome of the instinctive endowment of the child; but the critical question raised by Wellisch's thesis is whether these motives are not largely engendered by the environmental influences of the family situation.

W. RONALD D. FAIRBAIRN

The Leaven of Love. By IZETTE DE FOREST. London: Gollancz. 1954. Pp. xv + 206. 15*s.*

This is a simple and well-written book for the non-technical reader. It deals with 'The relation of the psychotherapeutic endeavour to the two Great Commandments, as exemplified in the theory and technique of Ferenczi'. 'Giving to the neurotic patient the hope and the healing experience of love makes way for a mutually loving relationship', the author points out on p. 19. 'This train of emotional circumstances in psychotherapeutic treatment has not as yet been adequately acknowledged or valued by the various schools of psychoanalysis. Yet for Sandor Ferenczi', she says, 'this was the sure measure of successful psychoanalytic therapy, the constant aim of his lifelong research.'

R. W. PICKFORD

Insanity, Art, and Culture. By FRANCIS REITMAN. Bristol: John Wright and Sons Ltd. 1954. Pp. 111. Price 12*s.* 6*d.*

This book may be viewed as a sequel to the same author's *Psychotic Art*, and embodies certain interesting researches which he has carried out on artistic work of mental patients, and a consideration of the relationships between their work and the cultural background. It has six illustrations and a frontispiece. The scope of the book may be gathered from a summary of the contents of its chapters. The first deals in a general way with the problems of psychotic art. The second deals with cultural influences, and after an examination of examples of psychotic art from a number of different cultural settings, suggests that such differences as are found in the art of psychotic patients in various cultures are due to cultural influences rather than to particular disease processes. The third chapter deals with the influence of the medium, namely that of embroidery or sculpture, used by patients, and it suggests that the use of the medium of sculpture tends to prevent the appearance of psychotic characteristics. The fourth chapter deals with an artist, George Morland, who was an alcoholic of manic-depressive temperament. It shows clearly that the art of such a man may remain unaffected by his psychotic tendencies. The fifth deals with psychic art, namely, automatic painting, doodles and spiritualistic products. Some features of these works are common to them and psychotic art, but such psychic art works are not obviously psychotic. They may be intermediate between psychotic and normal art. The sixth chapter deals with eccentricity and culture, and it discusses a well-known eccentric of the eighteenth century, John Mytton, a man of psychopathic personality, whose peculiarities were nevertheless socially acceptable in his time. In this chapter the author claims to show that unusual art may be acceptable culturally just like unusual behaviour. The last chapter is a summary.

The book suggests many interesting lines of research which might be followed out profitably concerning the relationships between culture, art and mental abnormalities. The author does not disguise the fact that all the problems dealt with in the book deserve much more detailed study. He has shown how material might be collected and handled and how statistical tests of significance might be applied to the differences of frequency observed. One feels that he has viewed a vast group of problems and applied various probes to the mass, and therefore his work is suggestive rather than conclusive, but it is unfortunate that he precluded himself from using the help which psychoanalysis might have given, by assuming that it is unscientific.

R. W. PICKFORD

Class Status and Power. A Reader in Social Stratification. Edited by REINHARD BENDIX and SEYMOUR MARTIN LIPSET. London: Routledge and Kegan Paul Ltd. 1954. Pp. 723. 42*s.* net.

Pages 17–45 of this volume are pure gold. They consist of reprints from classical writers, and it is delightful to learn from Aristotle that the value of the middle class resides in their ability both to command and to obey, to read Adam Smith pointing out the defects of capitalism in his beautiful, terse English, to discover that Marx believed people sometimes to act from motives other than love of gain, and to become acquainted with Veblen's surprising opinion about the value of good manners to their possessor. After that the book becomes verbose and uneven. Irritated by the lack of plan, the reader begins to wish either that Bendix and Lipset had written a book themselves, stating a few important hypotheses about social class, marshalling arguments for and against them, and quoting other writers sparingly, or, if they wished to make an anthology, that they had made a high quality of writing the test for inclusion.

One of the best contributions is Henry Pirenne's history of capitalism from the early Middle Ages to the present day, for, without hiding the complexity of the facts, he does seek to relate them in an orderly manner to a political and ethical thesis. Boris Brutzkus's outline of the economic history of Russia has a breadth of feeling that makes the reader sympathize with both the Tsarist régime and the revolutionaries, because he comes to understand the greatness of the difficulties with which they were battling. David Riesman propounds the interesting view that in present-day America neither generals nor trade union leaders behave as ruthlessly as their real power permits, because they are inhibited by inner psychological restraints, but he leaves his problem unresolved, because he has neither any clear opinion as to whether ruthlessness is good or bad, nor any awareness of the deeper motives underlying behaviour. In fact, practically nothing is anywhere explained in terms of the instinctual springs of action. There are, however, many thoughtful and informative essays, and a great quantity of statistics about various aspects of social class in the U.S.A.

JANE DARROCH

Motivation and Morale in Industry. By MORRIS S. VITELES. London: Staples. 1954. Pp. 510. Price 50*s*.

This must be the fruit of years of work. An enormous amount of sheer labour has gone into the task of compiling the results of hundreds of investigations; a remarkably clear mind has been at work collating it all and trying to turn it into a coherent whole. (Some of the chapters would make excellent models for the post-graduate student who is having difficulty with the section of his thesis headed 'Review of the Pertinent Literature'.) The smallest number of references quoted and listed in any chapter is 18, the largest 168. There are over 700 listed in the 'Index of Names' and 18 are British—which must be quite a record for American psychological works.

The book has five parts: I, Mobilizing the will to work; II, The background of motivational theory; III, Experimental studies; IV, Employee-attitude surveys; V, Guide posts for management. Part I is mainly concerned with setting and illustrating 'the problem'. Part II gives an admirably succinct account of relevant theories—incidentally putting Lewin and his associates and descendants clearly in the lead. Part III is a *tour de force* of compression. Part IV is not quite so good but uses many valuable sources other than psychological journals and is very informative. Part V, to my mind, makes a rather weak ending to the whole thing.

It was Part V which started me wondering whether I was reviewing *a book* or a work of reference. The preface states that it is meant primarily for management and only secondarily 'for courses in industrial management, personnel administration, industrial engineering and the like, as well as in courses in the field of psychology'. I think the primary and secondary purposes will be reversed in Great Britain. It is almost compulsory reading for all who talk about or practise industrial psychology; I cannot, however, see many British managers getting through the 500 pages—nor do I think they would get a great deal out of it all if they did. There is, of course, a theme to the book; it is printed on the spine. But (*a*) neither the problem-setting Part I nor the theoretical Part II is more than thinly related to the rest of the work, and (*b*) the conclusions of Part V do not follow with any logical force from the preceding pages. Viteles cannot be criticized for this: the fault lies in industrial psychological research—or, perhaps, in the awkwardness of human beings whose actions and reactions cannot at present be charted in any tidy way.

DENIS MCMAHON

First Report of the Joint Committee on Human Relations in Industry. (March 1953 to March 1954.) Department of Scientific and Industrial Research and Medical Research Council. London: H.M.S.O. 1954. Pp. 16. Price 1*s*.

Before the post-war Committee on Industrial Productivity was disbanded it recommended that the work of its Human Factors Panel should be taken over jointly by the Department of Scientific and Industrial Research and the Medical Research Council. Accordingly, in May 1953 two joint committees were appointed, one on Individual Efficiency in Industry, the other on Human Relations in Industry. The task of the Human Relations committee is to examine current researches in this field and to draw attention to problems on which research might be undertaken or supported by the Department of Scientific and Industrial Research, the Medical Research Council, or other bodies. At present the researches being wholly or partly supported fall into five main areas: Incentives, Management Organization, Promotion and Training, Special Groups in Industry (for example, married women), Technological

Change, and each of these areas is the special responsibility of a corresponding subcommittee. No details are yet available of the progress made by the researches.

As an example of how researches which are being sponsored interlock, the projects under the 'Incentives' heading include a critical evaluation of previous researches; a study of the impact of incentive schemes on productivity in terms of output, labour turnover, absenteeism, taking account also of trade union attitudes and experience; a study of the historical, economic and cultural factors affecting the establishment of production norms determined not so much by technical considerations as by attitudes of workers and management; a factory study of the process of changing from one method of payment to another; and finally an exploratory study of incentives for executive grades in industry.

As it happens, in this first year many of the researches are at least partly supported by Conditional Aid Programme funds; special conditions apply in that researches assisted by these funds must be directly related to productivity and industrial efficiency, and must show promise too of reaching applicable results before the end of the Conditional Aid Programme in 1956—conditions that do not decrease the difficulty of the Committee's task. However, although Conditional Aid has dominated the first year of the Committee's work and advanced it very considerably, the primary aim of the Committee is still to build up a 'balanced and continuing programme of social science research related to the needs of industry' as a bed-rock. Clearly one of its own main problems will be to identify the important topics for research and to assist in making the best use of the results. In identifying topics the perceived needs of industry form an important, though not by any means the only, criterion; the Committee's function as a link between industry and the existing social science research institutions is thus a vital one. A balance has yet to be found, however, between 'projects to aid the assimilation by industry of available knowledge, short-term applied research, and studies of a more fundamental character'.

R. M. MCKENZIE

The Modern Family. By ROBERT F. WINCH. New York: Henry Holt and Co. 1952. Pp. xxi + 522. $3.90.

Winch's book is chiefly concerned with the modern American urban middle-class family; this is, however, set in perspective by describing other varieties of American families, both past and present, and even families belonging to non-American cultures, and by keeping in mind the changes that the chosen type of family is now undergoing. The first third of the book discusses the principal functions of the family; the second third, the relations between parent and infant, parent and child, parent and adolescent, and parent and adult, each of these getting a chapter to itself; the last third deals with courtship, love and marriage. This last section is probably the most original and interesting part of the book; it is certainly the most entertaining, especially the chapter on 'Romantic Love in American Life'. The author's aim was to discuss the family's influence on and share in the formation of personality. He has succeeded in maintaining a proper balance between the sociological and psychological aspects of his theme. The book is clearly written, well organized, and well documented, and can be recommended either as a survey for the general reader or as an introduction from which the student can readily pass to a closer study of special topics.

H. E. O. JAMES

From Loneliness to Fellowship. By WILHELM AAREK. London: Allen and Unwin. 1954. Pp. vi + 79. 5*s*. net.

This little book is the published form of this year's Swarthmore Lecture, which is delivered at the beginning of each Yearly Meeting of the Society of Friends. The author is a Norwegian psychologist whose lecture is a thoughtful essay on man's need for fellowship and the means of achieving it. This has led him to discuss such topics as social integration, isolation, discipline and freedom, conscience, participation and so on, which he has done from a psychological point of view. But, oddly enough, it is the final chapter on 'Quaker Fellowship' which is most likely to arrest the attention of a psychologist, although this chapter ostensibly treats the problem from a religious standpoint. His account of the functioning of a Quaker Meeting, whether it be for worship or business, is worth some study by any psychologist who is interested in the dynamics of group life.

RALPH HETHERINGTON

LEARNING AND RETAINING VERBAL MATERIAL

By HARRY KAY

Institute of Experimental Psychology, University of Oxford

I. *Introduction* (p. 81). II. *Aims of the experiment* (pp. 81–82). III. *Experimental procedure* (p. 82). IV. *Subjects* (p. 83). V. *Treatment of data* (pp. 83–85). VI. *Results* (pp. 85–92). VII. *Discussion* (pp. 92–98): (i) *A Comparison of measures. Content: verbal and qualitative analysis* (pp. 92–93); (ii) *The relation of one reproduction to another* (pp. 93–94); (iii) *Relation of first reproduction to the original* (pp. 94–95); (iv) *The accuracy of initial recall—a comparison with Bartlett's findings* (pp. 95–97); (v) *The importance of initial responses in learning* (pp. 97–98). VIII. *Summary* (pp. 98–99). *Appendices* A *and* B (pp. 99–100). *References* (p. 100).

I. INTRODUCTION

In a series of experiments studying the formation and holding together of behavioural sequences (Kay, 1951, 1954), one of the major difficulties experienced by adult subjects was that of amending the mistakes which they themselves had introduced into their learning, so much so that in many instances the feature of the whole learning process was this unlearning of wrongly acquired responses. The following experiment was one of the attempts to examine the generality of these results by seeing if they applied in the verbal field, the most widely used of human communication systems.

Both Bartlett's (1932) technique and findings have considerably influenced the present study and, in particular, his observations upon the inaccuracy of reproduction and the persistency of a 'general form' once a first version had been given (*op. cit.* p. 93). The writer's previous experiments had confirmed this rapid establishment of a stable form which, though often far from accurate, was persistently maintained. In studying remembering Bartlett was not concerned with the difficulty of amending the errors in the 'schematized' version; in studying learning this was, for the writer, a major interest, and this experiment was designed to throw some light on the problem.

The question which was put then was in the form, 'Why should the initial reproduction be so stable when that version itself is an inaccurate reproduction of the original?' or alternatively, 'Why can accuracy of reproduction be achieved between one inaccurate version and another and not between the original and its first reproduction?'

There are many answers to this, and many of them have been given by Bartlett, but the question can be reopened by examining in some detail how far the first version is and is not an accurate reproduction of the original. Such an inquiry will not consider again the role of individual interests, attitudes, etc., in determining which parts of the original shall be accepted or ignored. Rather it will attempt to be more exact, within the limits of its own material, about the relationship between the original and the first reproduction.

II. AIMS OF THE EXPERIMENT

One aim then of the present experiment was to make a detailed investigation of the first reproduction, and consider its relation to the original. The design for this part of the experiment was straightforward enough: prose passages were read over to subjects, and after a few minutes they were asked to reproduce them in writing.

This brought the experiment to the stage where it could then study the second issue already mentioned—in what ways were subjects in their subsequent learning amending the errors in their first reproductions. The slight change from customary procedures of learning experiments which this involved will now be described.

III. EXPERIMENTAL PROCEDURE

Two prose passages were each read out twice to a group of subjects. Five minutes later, after listening to further material, subjects were asked to reproduce from the first *two* passages the one which they thought they could best remember. When they had finished writing this, they were asked to reproduce the non-preferred passage. After completing both, the reproductions were collected, and then the two passages were again read, this time once only. This marked the end of the procedure for that week.

One week later they were again asked to reproduce both passages and after doing this, the experimental session once more ended with the originals being read once to them. Another week later this routine was again followed and the experiment was conducted on these lines until six reproductions in all had been made. One slight departure had to be made—a fortnight, not a week, unavoidably intervened between the third and fourth versions. After the sixth version the students were told that was the end of the repeated reproduction series; the vacation followed and, at the end of the next term, four months from their last (sixth) effort, they were asked if they would attempt another reproduction, the seventh. This terminated the experiment.

The procedure has affinities with both memory and learning experiments. As with memory studies, there is some time elapsing before a subject is asked for his reproduction; as with learning studies, he is again presented with the original after his own version. The aims in compromising between the two types of experiment by extending the time interval between the presentation of the original and the actual reproduction of it were:

(1) On the one hand, to make conditions far from optimal that one reproduction should agree with another. (Both the long time interval and the interposing of the original were against this.)

(2) On the other, to weaken the effects of the original upon subsequent versions. Obviously the reproduction should follow much closer in time to the presentation of the original if the two were to agree.

Thus the wide time gap should, as compared with the normal interval in learning experiments, increase the inaccuracy both between one version and the next, and between one version and the original. These conditions for slow learning should facilitate the study of the relative influences and strength of the two versions—the original and the reproduction—upon subsequent reproductions.

The method, which is a slight variation on normal procedures, may be novel to experimental work, but it is certainly not novel to everyday life, for it is nearly identical with the school practice of giving instruction in a subject, then making some examination of what has been learned, correcting it, followed by further testing at some subsequent date.

IV. SUBJECTS

The subjects were attending a class in psychology at the Cambridgeshire Technical College (in 1950). The results of ten subjects, six men and four women aged between 18 and 35 years, are treated here. These subjects were able to complete six repeated reproductions and a final version four months later. As would be expected, an experiment (and a course in psychology) stretching over six months lost some of its starters, whilst many were not able to attend every session and therefore had gaps in their successive reproductions. Altogether results were collected from twenty-five subjects—the other fifteen have not been included in the analysed results because they have not seven successive reproductions. This was the only basis for their rejection, but a careful inspection of their records revealed the same trends as were found with the present experimental group.

When the class began, there were in it several American service personnel. The experiment was therefore planned with British and American subjects in mind, one of the two passages being in 'American-style journalism'. Owing to changes in personnel, none of the American subjects was able to complete all seven reproductions: hence the analysed results are all from British subjects and the original intention of making a detailed comparison between subjects reproducing a familiar and unfamiliar prose style had to be abandoned. Some of the broader differences between the versions of the two nationalities are briefly indicated in the results.

Material

The two selected passages were taken from Mr G. C. Grindley's Practical Class at Cambridge. One with the title 'Tea Leaves' (passage 'A') was an everyday piece of English prose, slightly and slyly humorous, describing the difficulties of selling tea. It contained 167 words, and was read first.

The second passage ('B'), entitled 'The Stenographer' was supposed to be a brash American advertisement, with short and sudden dramatic statements. It aimed to be funny and 'smart'. It contained 134 words.

Both passages are reproduced as they were read to subjects in Appendix A.

V. TREATMENT OF DATA

Verbal analysis

There are many difficulties in trying to compare one prose passage against another. As with a work of art, certain broad divisions may be generally approved but the more subtle and individual characteristics—particularly those of style—find no unanimous acceptance. It is therefore not unexpected that quantitative and qualitative evaluations of prose reproductions have often been irreconcilable. The experimenter is faced at one extreme with the problem of collecting a mass of quantitative data of which he feels, or should feel, considerable doubt, and at the other with the recording of his subjective impressions with little chance to indicate their generality and frequency.

In this experiment, for example, one subject in trying to reproduce the phrase 'the door of a speakeasy' gave it for four consecutive weeks, in spite of hearing the correct original after each version, as 'the door of a speak-quickly'. The effect on a reader, however, may be disproportionate to its frequency. On the other hand, the obvious quantitative units—sentences, phrases, words or even syllables—are not interrelated on a scale of measurement, and the choice of any one may in part determine the result.

It would seem that there is no one solution to this problem but that the answer is dependent upon the length, style, subject-matter, etc., of a passage. The treatment of this experimental material was empirically determined from pilot studies with a wide range of passages. Scoring was based on word counts; there were four categories:

Correct—words reproduced accurately.

Omissions—words in the original not reproduced.

Errors—the committed errors: words substituted in the reproduction for those in the original.

Additions—words in a reproduction for which there was no corresponding word in the original.

(A further category of 'misplaced words' was considered and found unnecessary with this length of passage.)

It was felt that the four categories did provide a fair overall impression of the accuracy of a particular version. The word count is not ideal since it disregards the interrelated character of prose and treats all units as if they were of equal difficulty and significance. But these or similar objections may be laid against other possible measures, particularly the popular method based on phrase divisions, whilst the word count has the advantage of being the most objective procedure.

Content analysis

For this second measure, which was independent of the word counts, the experimenter split the content of the passages into ten sections. Three independent judges did the same, and agreement was reached on the ten major divisions. (These are given in Appendix B.) Each reproduction was then examined to see how far and in what ways its content did comply with that of the original. For each of the ten sections a subject was scored either right, wrong, or omitted. The exact fidelity of the vocabulary was not important here.

Repetition analysis

The verbal and content scores were used both in considering the accuracy of the first reproduction and in clarifying the second major issue—the difficulty of amending errors in subsequent reproductions. In a further effort to throw light on this difficulty several counts were made on the number of repetitions in a reproduction. These included for any one subject how far any one version was repeating the words and content of the *immediately* preceding reproduction, and also how far such words and content had been repeated in *all* preceding productions. Details of how these analyses were carried out will be given in the section on results.

Qualitative findings

It was intended that the objective evidence from the content and verbal analysis should be considered in conjunction with more qualitative judgements, which it was hoped they might substantiate. The qualitative treatment of the data was on traditional lines but one further examination was conducted. The seven reproductions of each subject were typed out and given to an independent judge. Then the ten reproductions for each week were randomized, and the judge was asked to assign them to their respective authors, on the basis that there was some similarity between the successive reproductions of the same subject. This was done for all six sets of reproductions—the second to the seventh inclusive—making a total of sixty choices in all for each judge. The selections and the reasons for them were recorded from the judges.

VI. RESULTS

Verbal analysis

The data based on the scores for all subjects for both passages are set out in Tables 1A and B. The figures show the mean number of words in each reproduction and the analysis of this total into 'correct', 'errors', 'additions' and 'omissions'. The standard deviations are small enough for the mean scores to be taken as indicating the general trend.

Table 1 A. *Verbal analysis of passage A*

(Mean number of words reproduced in each version and the mean number repeated from one reproduction to the next.)

Versions ...	1	2		3		4		5		6		7	
	M	M	R	M	R	M	R	M	R	M	R	M	R
Correct													
Mean	44	44	32·3	53	35·6	56	45·8	62	49·8	67	56·0	53·8	47·0
S.D.	11·3	13·8	12·8	18·5	15·1	16·7	19·9	16·4	18·5	17·4	17·6	21·5	18·8
Errors													
Mean	48	48	25·2	44	27·9	45	32·4	43	31·9	40	31·0	46·2	30·1
S.D.	8·4	5·8	5·5	7·5	5·0	3·6	4·2	6·0	5·0	8·2	7·0	5·5	7·4
Additions													
Mean	21	17	6·1	22	6·5	17	8·4	16	7·0	16	7·0	21·2	8·1
S.D.	14·6	11·9	3·4	12·1	4·4	8·4	5·8	5·2	8·7	7·1	3·8	9·4	6·6
Omissions													
Mean	74	75	52·6	69	53·7	66	54·1	62	53·0	60	52·7	65·7	52·8
S.D.	11·6	13·6	12·0	14·3	7·7	15·6	7·5	15·8	12·9	17·6	14·5	19·8	16·2
Total words													
Mean	113	109	63·8	120	70·0	118	86·6	120	88·7	123	93·9	121·0	85·2
S.D.	22·5	22·4	13·3	14·4	13·6	21·3	17·9	14·4	18·5	18·9	16·5	14·8	16·7

Columns M give the mean number of words reproduced in each version, columns R give the mean number of words repeated from the immediately preceding version.

Table 1 B. *Verbal analysis of passage B*

(Mean number of words reproduced in each version and the mean number repeated from one reproduction to the next.)

Versions ...	1	2		3		4		5		6		7	
	M	M	R	M	R	M	R	M	R	M	R	M	R
Correct													
Mean	54	52	42·1	59	46·2	55	50·7	66	52·8	71	63·2	57	53·9
S.D.	9·6	14·6	11·4	18·8	17·8	14·8	17·6	11·5	15·2	12·1	12·2	14·6	13·3
Errors													
Mean	37	38	27·2	35	26·8	37	29·8	35	29·3	32	28·8	35	27·8
S.D.	10·0	8·6	5·9	10·0	8·5	9·3	8·2	9·1	8·4	8·5	8·5	7·6	6·3
Additions													
Mean	20	20	8·3	22	8·7	17	8·1	21	9·2	17	11·1	15	8·1
S.D.	8·1	8·1	3·9	7·6	2·9	6·8	3·2	8·4	3·4	5·7	3·9	5·3	4·0
Omissions													
Mean	42	42	29·2	40	33·2	42	30·7	35	31·8	31	30·0	41	27·8
S.D.	10·0	9·0	5·3	8·1	7·7	10·0	4·9	4·4	3·7	5·6	5·0	7·0	4·6
Total words													
Mean	112	112	77·6	116	81·7	109	88·6	120	91·3	120	103·1	108	89·8
S.D.	15·4	8·2	13·4	15·8	17·5	12·4	13·8	7·1	8·4	5·9	6·1	5·3	6·2

Columns M give the mean number of words reproduced in each version, columns R give the mean number of words repeated from the immediately preceding version.

It will be seen that the mean total number of words did not substantially change, the increase for both passages throughout all the reproductions being less than 10%. The mean number of words correctly reproduced was slowly rising until in passage 'A' it represented a 50% and in 'B' a 30% increase. But this increase was not at the expense of any one of the other three categories, which all showed some decrease. This can be seen clearly from the histograms in Fig. 1. If for each reproduction the categories are ranked according to their magnitude the only difference between the rankings for the first and sixth versions is that the 'corrects' now occupy a higher position; the others retain their rank orders with surprisingly little change in their relative magnitudes.

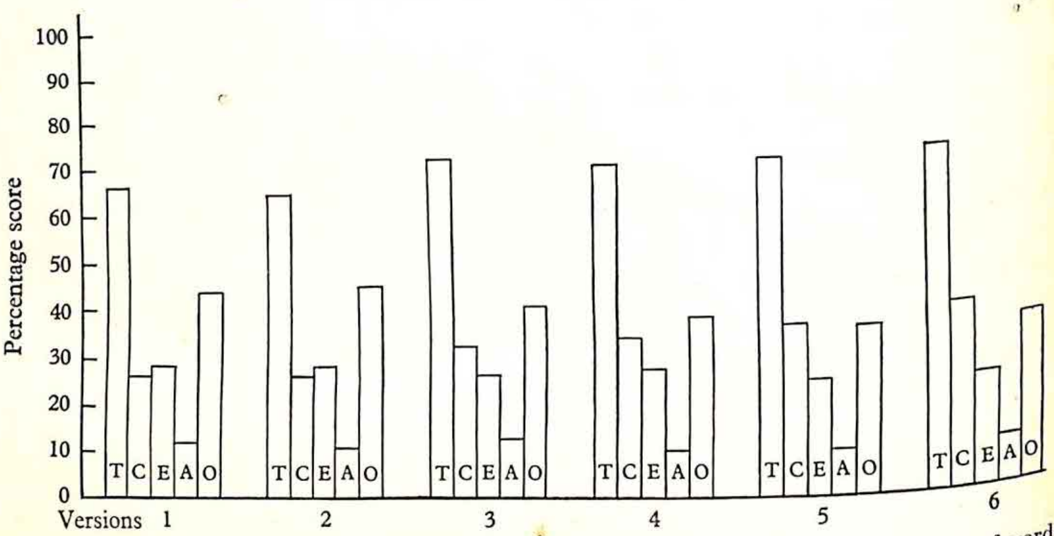


Fig. 1. Verbal analysis of the six reproductions. Passage A, percentage scores; T, total number of words; C, correct number of words; E, errors; A, additions; O, omissions.

A popular idea of learning would be that mistakes are righted, but it would seem from the above analysis that the accuracy of the passage improved partly because there were fewer omissions, partly because there were slightly fewer additions and partly because some errors were rectified. The relative improvement in each category was slight compared both with the overall improvement, and with the still remaining mistakes.

Since the passages were of different lengths—167 and 134 words—Table 2 compares the verbal scores by expressing the results for each version as a percentage of the actual number of words in that passage. The table brings out clearly that relatively more of the words of passage 'B' were being correctly reproduced and that there were also fewer omissions; to this extent the verbal findings accord with those for the content of the two passages.

Verbal repetition

The number of mistakes which occurred at each reproduction made it probable that many were being repeated from one version to the next. A measurement of this repetitive tendency has been attempted in the 'R' columns of Tables 1 A and B. This gives for each reproduction (in columns 'R') the number of words which have been repeated from the immediately preceding version. If, for example, in passage 'A' we consider the 123 total words reproduced in version 6, 94 of these were repeated from version 5. Of these 94 words, 56 out of the 67 correct words were already reproduced in version 5, as were also 31 out of the 40 errors and 7 out of the 16 additions. Even this does not end the 'repeat story'; there were 60 omissions in version 6 and of them 53 were also omitted in the previous reproduction. It is fairly obvious that one version was bearing a much greater semblance to its immediate predecessor than either was bearing to the original.

Table 2. *Comparison of the verbal scores in the two passages*

(The means of Table 1 expressed as percentages of the actual number of words in each passage.)

Version	1	2	3	4	5	6	7
Correct							
Passage A	26	26	32	34	37	40	33
Passage B	40	39	44	41	49	53	42
Errors							
Passage A	28	28	26	27	25	24	27
Passage B	28	29	26	28	26	24	26
Additions							
Passage A	12	10	13	10	10	10	12
Passage B	15	15	17	13	16	13	11
Omissions							
Passage A	44	45	41	39	37	36	39
Passage B	31	31	30	31	26	23	30
Total words reproduced							
Passage A	66	64	72	71	72	74	72
Passage B	83	83	87	81	90	90	80

Table 3. *Verbal repetition from all preceding reproductions*

Versions ...	1		2		3		4		5		6		7	
Passage A	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%
Correct	—	—	32	73	30	93	29	97	28·5	98	28	98	27	96
Errors	—	—	25	52	21	84	18·5	88	17	91	16	94	15	93
Additions	—	—	6	35	4	66	3	75	2·5	83	2	80	2	100
Omissions	—	—	52·5	71	47·5	90	45·5	95	44	96	43	97	40	93
Passage B														
Correct	—	—	42	77	40	95	39	97	38·5	98	38	98	37	97
Errors	—	—	27	73	23	85	22	95	21	95	20·5	97	19	92
Additions	—	—	8	40	6	75	4·5	75	3·5	77	3	85	2·5	83
Omissions	—	—	29	69	26	90	24	92	23·5	97	23	97	22	95

Each score gives the mean number of words in that reproduction repeated from *all* preceding reproductions. Percentage scores express the above mean scores as percentages of the total possible scores for that reproduction.

Table 1 has illustrated the prevalence of a certain kind of word repetition—that from the immediately preceding reproduction—but it now seemed relevant to ask how frequently was a word repeated when it was correct and when it was wrong. There was the further question of the probability after a mistake had *often* occurred of it being repeated, and whether it was more likely to be reproduced again than a mistake which had been made only once. Table 3 attempts to answer some of these points; it details the mean number of words in each version which have been reproduced in *all* the preceding versions. Thus if we are considering the errors in version 5 (passage B) the table is only concerned with the 22 errors which have already been reproduced in versions 1–4: of these 22 errors 21 were again made in version 5.

This type of measure is of course very strict, for if a word drops out only once in any reproduction it is thereafter lost in all the measurements of subsequent versions. It will be appreciated that this implies that in Table 3 in any one category the total possible score for one reproduction is the actual score of the immediately preceding version; and similarly that score, by the same definition, could not be greater than that of any of its preceding versions. Hence almost inevitably there is some decrease in the means from one version to the next. But of the words which do comply with these requirements a very high proportion are subsequently reproduced. This is illustrated in those columns of Table 3 which express the results as a percentage of the total possible score for that reproduction. In the above example, 21 errors were repeated of a possible 22, that is 95%.

The results expressed in the table seem definite enough—if a word was reproduced three or four times then the probability that it would be reproduced in a subsequent version was extremely high. In this experiment where learning conditions were slow that probability was nearly as high for an error as for a correct word. The point also applied to words appearing in the original and which had been repeatedly omitted—there was a high percentage of words which subjects had never attempted to reproduce and the longer such words had been ignored the more likely in fact was this to happen.

It should be pointed out that the 'Additions' which were repeated, represented only a very small proportion of the total additions for most of the reproductions. The words which a subject added to a passage, seemed to vary more from one reproduction to another than do words in the other three categories—corrects, errors, and omissions. Hence, though the percentage figures for the additions were high, they were percentages of very small totals and are not reliable.

Content analysis

As already indicated, all the reproductions of every subject were examined to see how far the content agreed with that of the original.

Table 4 gives the results. If we begin by considering only the content which was actually reproduced at each version it is obvious that a very high percentage of both passages could be recalled at the first attempt. Within whatever individual framework subjects were holding and expressing their ideas there is no doubt that most of the substance of each passage had been assimilated. The striking feature *is* that so much had been grasped from the beginning.

Table 4. *Total scores for each reproduction of content of passages A and B*

(Total possible score for each reproduction—10 items × 10 subjects = 100. Correct, C; omissions, O; errors, E.)

Versions ...	1			2			3			4			5			6			7		
	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C	O	E
Passage A																					
Content actually reproduced	59	28	13	51	38	11	61	28	11	64	25	11	63	25	12	68	23	9	62	25	13
Content repeated from immediately preceding version	—	—	—	51	28	10	51	26	9	60	25	10	63	25	11	63	23	9	61	23	8
Passage B																					
Content actually reproduced	83	13	4	86	12	2	86	12	2	83	12	5	86	9	5	90	6	4	83	11	6
Content repeated from immediately preceding version	—	—	—	81	11	2	83	11	2	82	11	2	82	9	4	85	6	4	82	10	4

It would be suspected from these results that the content actually reproduced in one version would be very similar to that of another. This further analysis was carried out, and Table 4 also reproduces the scores showing how far the content of one reproduction repeated that of the immediately preceding version. The figures do establish very definitely how little change there was in content from one version to the next. As some learning did take place there was a slight trend for more substance to be correct and less to be omitted, but the newly acquired substance was added to that which was already held whilst that which was finally omitted had never appeared.

It had been intended to establish this point by a further analysis of the content, examining how far a reproduction was repeating the substance of *all* preceding versions. But it was felt that the data of Table 4 sufficiently indicated that so far as content was concerned, subjects were repeating the same correct items, making the same errors, and omitting the same items one week after another. In other words, the initial reproduction had established the content of the passage; as it was laid down then so, with only slight modifications, was it held six weeks and ultimately six months later.

Qualitative findings

We can now turn to the more qualitative analysis of the passages. In order to obtain some indication of the individuality of a reproduction all the versions of all the subjects were typed out on separate sheets of paper; following the procedure already described (see p. 84) five independent judges were asked to select at each successive stage the six reproductions of each subject. No mistakes were made, and it was obvious that though the process was tedious there were ample cues upon which to base the assessment. (The statistical significance of the result even with five judges is of course very high.)

When the judges were asked to say upon what they had based their judgements, in addition to such points as the amount of the passage reproduced, the particular parts omitted, and the general accuracy of the ideas and the words, they also stressed the recurrence of key words and phrases, and the ease with which certain individual stylistic features could be recognized. It is hoped to illustrate some of these features in the following examples.

Minor but persistent verbal changes

In this type of example the verbal alterations were of a very minor order incurring little or no change in meaning. This often makes it the more interesting because such 'error' words could have been so easily interchanged. It was also noticeable that they occurred in sections where much was accurately reproduced. A typical example was where a subject, no. 1, in all his seven versions wrote for 'the actual political opinions of mankind' the phrase 'the political ideas of man'. As a change in meaning it is negligible but as a change that can persist over six months it is of note. Similarly, the same subject, again in all his seven reproductions, uses 'the entire stock' for 'the whole stock', a 'vague urge' for 'a vague impulse' (the word 'urge' was frequently used by other subjects in this context) and 'a picture of the Tropics' for a 'vague suggestion of the Tropical East'. Another subject, no. 3, who had been in the East, in his seven reproductions always referred to the tea as growing 'on a bush' and in three of these it was 'on a bush in Ceylon'. Subject no. 7, a grocer by trade, instead of the phrase 'What therefore should the advertiser do to create a commercial entity', always gave 'What is the *purveyor* of tea to do to bring his product'. For the same subject the 'subconscious memory of a geography lesson' was always 'a half forgotten geography lesson'.

The same persistence in adhering to minor adjustments of the text was found in the reproduction of passage B. The phrase 'instead of throwing tear-gas bombs' caused many errors; one subject, no. 8, in all his seven versions gave 'faster than tear gas' whilst subject no. 9, apparently economically minded, offered in his first five versions 'saving the police using tear gas' and then in his last two versions extended this to 'saving the police the expense of using tear gas'.

In the last sentence of the passage the phrase 'to have the office heavily insured' was usually reproduced more elaborately, as in the six repeated versions 'should ascertain he has a full fire insurance'. Several subjects apparently seemed to think they were catching the American idiom by following the popular principle of never saying in one syllable what could be said in two. It may have been that certain phrases seemed to possess some American quality which influenced subjects in adhering so repeatedly to them: no. 2 maintained his incorrect 'city police' for five reproductions only to change it in favour of the more inaccurate 'chief of the city police'; subject no. 5 was equally insistent on repeating 'steel framed equipment'.

As was observed at the beginning of this section, it is not the magnitude but the persistence of such deviation which is the relevant point. It does seem surprising that at weekly intervals, and sometimes longer, anyone should repeat so exactly what they have previously reproduced when their last perception was not of their reproduction but of the original which corrected their errors.

Some major verbal changes

The examples in this section were often more noticeable than the types so far discussed, and perhaps for this reason they were somewhat less persistent.

Some of the changes, without introducing any different meaning, were striking in their new phraseology as when one subject (no. 8) in six consecutive reproductions referred to 'a name with an Eastern flavour'. Subject no. 10 would persist for five reproductions in using the phrase 'man in the street' (which appears in neither original) and even included it twice in two reproductions. The phrase dropped out of the sixth reproduction but returned after the four months' interval in the seventh. An interesting embellishment in which several subjects indulged was to make a 'geography lesson' into a 'school geography lesson'. Here perhaps we are not dealing so much with a change in phraseology as a persistent way in thinking—namely that geography lessons are usually (or ought to be?) school lessons.

The type of alteration which did involve a change in meaning was that by subject no. 5 who amalgamated 'the geography lesson on Australia' with the phrase 'the recollection of his grandmother' to make for five versions 'the grandmother in Australia', whilst subject no. 3 perhaps influenced by popular notions of grandmothers and tea drinking—there used to be many pictures on tea caddies of old ladies holding teapots—three times reproduced 'the recollection of Granny's teapot'. The strength of this tendency to incorporate general ideas into specific material may be gauged by the habit persisting even under these conditions of reproducing and then presenting the original material at weekly time intervals; presumably this type of mistake would be more prevalent if the intervals were longer.

In passage B some of the omissions brought out how easily phrases were dropped out when they did not advance the narrative. The unnecessary 'into the building' was continuously omitted by nearly all subjects, and the description 'in front of the door' was rendered by the majority as 'outside the door'. It was perhaps because of a lack of familiarity with the idea of a police 'raiding squad' that everyone referred to 'the police' but many omitted 'raiding squad'. It was also noticeable that most subjects began by saying 'I type quicker than a new band', etc., and omitted any reference to 'take shorthand notes'; it was as if typing had been the main subject-matter of the passage and had pushed out any reference to the other quality in spite of it occurring at the beginning of the original. There was the usual persistence on a number preference, one subject insisting on interpreting 'several large firms' as 'the last three firms' and in two versions she also 'set fire within three days' to the offices, whilst another subject would repeat 'three weeks' for 'a few days'.

The effort to reproduce 'Style'

So far examples have quoted short phrases and brought out alterations in individual words. And even when illustrating qualitative results it is difficult to do more than this. But reading through the reproductions it was obvious that a subject had reacted to much more than we have so far conveyed. The words were merely the hard structure of each creation; within this verbal framework not only had a set of ideas—the main content of the passage—been more or less accurately conveyed but they had been transmitted with very obvious awareness of the stylistic features of their presentation. The reactions to the different styles of the two passages was most emphasized. At the same time the attempts to reproduce that style were often anything but successful, and as already indicated, the styles of any one subject's reproductions tended to bear more relation to each other than any one of them did to the original.

This point was most clearly brought out to the experimenter when comparing the versions of passage 'B' by the American and British subjects. Though in many ways the mistakes of both nationalities were the same, and along the lines already indicated, there were stylistic differences, notably that the American subjects had extended the ideas, and sometimes the jargon, of the original. (The American subjects without exception had chosen to reproduce 'B' before, and at greater length, than passage 'A'—the British had done the reverse.) As examples of their reproductions, the bootleggers became 'a couple of Prohibitionist bootleggers' for one subject in all his four versions; another maintained it was the 'Oklahoma City Police Vice Squad'; a third made it 'F.B.I. officials'; a fourth described it as 'the Police riot squad'; whilst a fifth related how 'the State Police hired me to sit on the side walk'. The American idiom is obvious enough here as it was with the only subject who put the passage into the third person and began 'A small town in Oklahoma. . .' and finished all four versions with 'She warned her next employer to watch his step as he was going to get quite a girl'. (The three American subjects who gave any indication on this point all assumed the stenographer to be feminine.)

As a very brief summary of the qualitative differences between the versions of the American and British subjects in this passage, it may be said that the American reproductions were elaborated, almost casually and inevitably, by a familiarity with a wealth of background material. When a subject moved outside the limits of the passage he added local colour to it. It would be difficult to distinguish how far content and how far particular idioms and phrasing were contributing to this total impression. But the result was that many of their versions had lost the pseudo style of the original and sounded authentic. British subjects, on the other hand, when they departed from their text, were often too obviously striving to achieve a style with the result that their versions were more uneven. They could be consistent with themselves, but not with the original.

The total impression given by the reproductions of the two nationalities was very different, but their versions established clearly enough what were the main reactions to the passage—a generally accurate construction of what it has said and at the same time a definite awareness of how it had said it. These together made up the passage, but matter and manner were modified against the background of the individual's experience, and the process seemed as recognizable under these conditions of learning as it was under the customary Bartlettian Remembering Studies.

VII. DISCUSSION

(i) *A comparison of measures. Content: verbal and qualitative analysis*

The relationship between the results of the content and verbal analysis has been graphed in Fig. 2. Whilst approximately 60–70% of the content was accurately reproduced in most versions, the percentages of correct words was only half of this, lying between 25 and 40%. Conversely, only a small percentage of the content was wrongly reproduced—9–13%—whilst the verbal errors for all reproductions was around 25%. The differences between the omissions are not so marked but are still definite enough—for the content they lie mainly between 25 and 28%, and for the verbal from 36 to 45%.

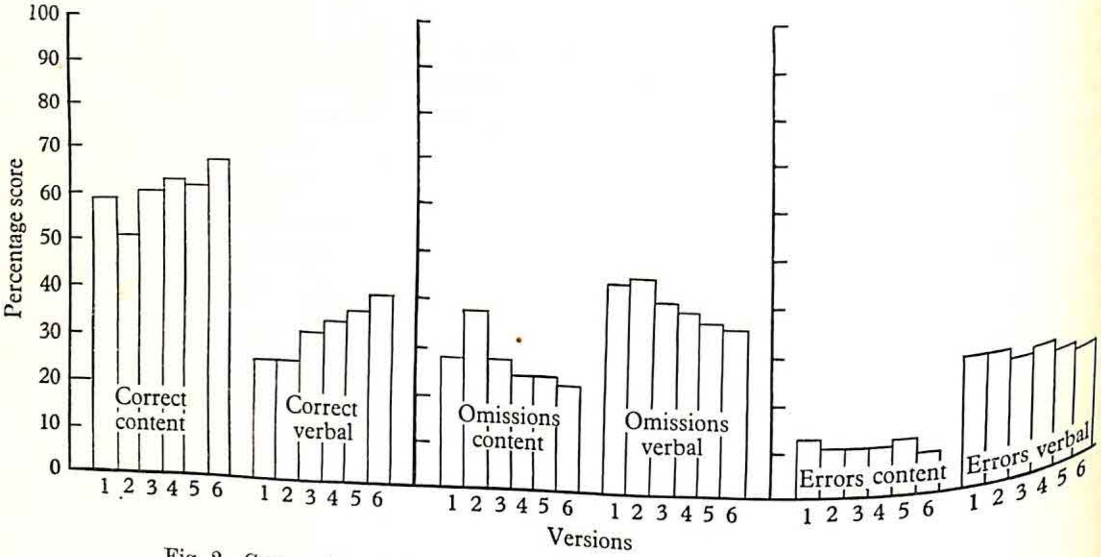


Fig. 2. Comparison of the content and verbal analysis. Percentage scores.

The discrepancy between the two scores raises the problem as to how the accuracy of a reproduction should be assessed. Here if only the content is considered, then it is a fairly satisfactory rendering of what has been said; but if the actual vocabulary of the original is required, then it is not.

It is fairly obvious that no inflexible rule can be laid down for making a judgement between these two criteria. If the passage is a short one, then the accuracy of the verbal reproduction will improve in comparison to the content analysis; with increasing length it will probably suffer. But this is supposing that content and vocabulary remain within one range of difficulty. Where, for example, the content becomes complex or abstract, then its reproduction is likely to suffer accordingly—some items in the 'A' passage which were consistently reproduced inadequately in the present series illustrated this point. On the basis of the present results it can only be argued that both measures should be considered if anything like an overall impression of a version is to be achieved.

In point of fact this has not often been done; it would follow that where results of different verbal experiments are being considered, it is essential to know the criterion upon which they are based. Generally more weight has been attached to the qualitative observations. It would seem almost inevitable that such observations would magnify differences between a reproduction and its original—where the discrepancies are striking, the reader tends to be over-influenced by them and does not allow sufficiently for the relative infrequency of such examples. For this reason the verbal analysis is providing useful supplementary evidence to the more qualitative demonstrations of verbal changes. The two may be regarded as complimentary—the analysis giving the frequency and the qualitative examples indicating the magnitude of such errors. In spite of the doubts which were expressed about the verbal analysis, it is felt that with this length of passage it can serve a useful function; at the same time it is the study of the words themselves which is likely to throw most light upon what a subject was doing, and trying to do.

(ii) *The relation of one reproduction to another*

Taking into consideration all the above measures, we can now take up the question of the relation of one reproduction to another. Though most studies of serial reproduction have stressed how closely one version follows another, in the present experiment, with the interjection of the original after each reproduction, there was added opportunity for subjects to amend their errors. In a very limited way they did this; they were learning, but the outstanding result was that whatever the measure considered, each subject established an extraordinarily close relation between one reproduction and another. This applied to the form and style of the passage, to individual words and minor items, in addition to the general content and overall impression. The result was, in effect, that where a subject had mistakenly introduced errors into his reproduction for two or three weeks, though he had heard the original with its corrections of his mistakes after each of his versions, the chances were very much in favour of him not amending those mistakes but of repeating them in exactly the same form. The point also applies to words which had been omitted—some words in the original were never reproduced, and the longer they were ignored, the more likely was this to happen.

Thus, the reproductions which a subject was making were steadily becoming more rigid and predictable, in that with each succeeding attempt a reproduction was becoming more and more like its predecessor. If the original was ever to be learnt, these errors of omission and commission had to be overcome, but it was obvious that this was not an easy step, in so far as certain words, phrases and ideas, had been 'built into' these early reproductions.

It would follow from what has been said that the first reproduction was all-important in laying down the future lines upon which a subject would approach the material. Thereafter, consistency was achieved in spite of the 'intrusion' of the original. This has an obvious bearing upon learning studies which will be illustrated in the last section; meanwhile, it leads us to a consideration of the issue—the relation of the first reproduction to the original.

(iii) *Relation of the first reproduction to the original*

In answering the question 'How accurate is the initial reproduction?' Bartlett was definite enough: 'Accurate recall is the exception and not the rule' (*op. cit.* p. 61; cf. also p. 93). Bartlett, of course, is not referring specifically to the initial or any other reproduction; but his schematic theory of memory makes it obvious that he would be regarding the initial reproduction as one stage in this continuous series of dynamic changes.

With our material, Bartlett's judgements would again apply to the initial reproductions; only about 30% of the actual words in the two passages are correctly given. But the statement, whilst accurate enough within its own limitations—'in a literal sense'—would be a misleading summary of the more general features of the reproductions. Most of the content—approximately 70%—was being accurately conveyed by the initial versions; on this criterion the first reproductions are satisfactory whilst their general form and phrasing reflected a definite effort, though not necessarily successful, to convey the different styles. Essentially these reproductions represented a total reaction to the original material—total in so far as it included not only a reaction to the obvious features such as the content, but also to the more subtle such as how a thing was being said, what kind of a person might have said it, what was its background, etc. In discussing the passages with subjects after the experiment it was frequently brought out that without having formulated their reactions they had some fairly persistent impressions which were more definite than they themselves realized; they had categorized material and events without necessarily being aware that they had done so. It was only when they were asked why they were making a particular statement that they realized their previous assumptions. This could be illustrated by precise examples—the air-crew pilot still engaged in flights over Germany who four times reproduced the 'varied political opinions of mankind' as 'varied as the political opinions expressed in *Mein Kampf*'; the subjects who had 'reproduced' a feminine stenographer, and those who had assumed a sex without stating it; but over and above these particular instances there were the more generalized impressions which seemed to have permeated and be effective in determining what a subject was attempting to reproduce.

This point was clearly illustrated in another context when using *picture* material with the method of serial reproductions. After subjects had heard a description of the picture and passed it on, they were questioned about their other impressions of it; what period of history did it represent? the relative sizes of objects? could they draw it? etc. The answers were generally negative and an impression was conveyed of extreme vagueness or even blankness. Then the original was shown to them and the instantaneous effect upon a subject was almost always one of incredulity—often a gasp of astonishment. When asked why, a typical naïve reply was 'I did not imagine it to be like that'. 'But you said you couldn't imagine it.' 'No, I couldn't, but somehow that is different.'

In other words, the subject was holding more than he could convey, and what he was holding was formed from a reaction between the limited new material he had been given and the almost unlimited resources upon which he could draw and against which he had assimilated the new. In the present passages, as with picture material, the subject was aware of what he had been given, but was unaware how far he had made a subjective reassessment of it. With the picture material this came out graphically when two subjects could agree upon a verbal description of certain details but could not agree upon the authenticity of the original picture; that is to say, they could agree upon what they themselves knew they were holding—their disagreement was arising over their holding so much more than they were able to express.

We may now ask how far these findings on the general accuracy of the initial reproduction would agree with Bartlett's exposition.

(iv) *The accuracy of initial recall—a comparison with Bartlett's findings*

Bartlett's monograph served magnificently, not only to illustrate the fallacies in everyday remembering, but to examine where and why they occurred. There is, however, the danger that the victory has been too complete. Having destroyed the myth of a verbatim accurate reproduction and seriously damaged a trace theory of memory in favour of a more loosely conceived and modifiable scheme, the question now arises: How far are memories ever accurately conveyed, and, if they are accurate, how is this achieved? Fortunately we are attempting to answer the first question, though this may throw some light on the second.

Bartlett's emphasis is always on the point that recall is a constructive process, largely determined by an attitude which is itself arising from the initial impressions left by the original (e.g. *op. cit.* pp. 176, 205, 207). 'It is this impression, rarely defined with much exactitude, which most readily persists', and the details are built into and around it so that they may satisfy the general setting.

This is, of course, a difficult point to discuss, since obviously we can only deal with such a process at second-hand through its products; here we can never say what that impression was, but, of necessity, have to base our assumption upon the results of some subsequent reproduction. And it may well be that under everyday conditions, where individuals are perceiving material without being directed by any instructions to remember it, their constructive reproduction is as imprecise and affectively determined as Bartlett has made it out to be. But under conditions such as the present, Bartlett's observations hardly seem to give sufficient weight to the accuracy with which the substance of the passage could be initially reproduced. This would apply to the results of all our subjects: in examining their first reproductions, we seem to have the product of some process whereby the material has not only been assimilated, but, to some extent has been carefully sifted and evaluated—the version which nearly all subjects reproduced did contain the main items in the passage.

The closer Bartlett's work is examined, the more would it seem that the examples which fit his account of constructive recall, and which he himself most frequently mentions, are those where subjects are trying to recall material after a long time interval—at least after some months and sometimes years. In some cases so much of what was originally given has all too obviously been 'lost'. The subject struggles to create something of the attendant circumstances—the time, place, people, etc.—anything upon which he can hinge a specific item and he is tremendously influenced in this searching by the probability of an event. The writer studied this process by asking student-subjects the question 'Can you recall your first University lecture?' and noted their introspection. It often became impossible for a subject to distinguish whether an event had taken place because he remembered it, or because he knew it was likely to have happened at that time. With the recall of events of this order Bartlett's exposition of a constructive effort seems apt enough, but there is a pronounced difference in degree, if not in kind, between this long-term remembering and the recall of material that has only been held for a few minutes. There is obviously less searching over a range of possibilities, certain items slip out promptly whilst others are quickly grouped around them. The question arises how far is this itself a constructive process, and how far is it the product of a 'constructive' process which has already taken place—in fact, which has taken place at the time of the initial perception?

Bartlett himself provides evidence on this point for it is of course an integral part of his thesis that perceiving and remembering have the same active characteristics, both being dominated by what he has called the 'effort after meaning'. But by reiterating the importance of a constructive process at the time of recall, Bartlett does not seem to have given sufficient weight to his own earlier observations of the active nature of perception itself. If his observations are right, as they seem to be, then at the first presentation of the material this effort to extract 'meaning' has begun. Bartlett has indicated the variables which influence the procedure, but we are concerned with the effect of this assimilation of material on this perceptual stage. To some extent passages may suffer an unwitting transformation, speech habits and verbal idiosyncrasies are incorporated into a new version according to personal predilections and social conventions. But the most significant feature is that this perceptual analysis—and in many ways this does seem to be a fair description of the process—is more accurate than is generally allowed. Whatever we may be conceiving as the retaining mechanism by which this is stored, it results in an appreciation of what has been presented, which, by overall standards, is a comprehensive digest of the material.

The argument then is stressing two points:

(1) That where material is being remembered after relatively short time intervals, it is necessary to give more weight to the interactive processes which have taken place on the initial perception of the material, and less to the constructive effort at the time of recall.

(2) In so far as the first stage of remembering—the initial perception—is an interaction directly involving the presented material, whereas the constructive stage of the final recall does not, the first is the more accurate. (It is assumed here that retention is a dynamic process, involving the possible modification of remembered data. Hence, a subject in trying to recall material is not necessarily holding the identical material which he began by retaining.) In view of the nature of this initial interaction, it seems somewhat misleading to speak of it as a 'constructive' process, even if we are prepared to apply this term to the recall stage of remembering. Accordingly, in an effort to place more emphasis upon the presented material it is proposed to describe this initial assimilation, not as a 'constructive process' but as a 'perceptual analysis'. It may well be that under particular conditions this analysis is defective—where, for example, the material is abstruse or excessively lengthy or disconnected so that items may easily be omitted. Under normal conditions it is not *literal* recall, but its deficiencies, though striking, should

not obscure the fact that it successfully assimilates a mass of material that would be well beyond the immediate memory span.

Since completing this experiment the writer has had the opportunity to read and discuss the research which Gomulicki, working under Zangwill, has done upon 'Recall as an Abstractive Process' (Gomulicki, 1952). As this is now being published the only comment which should be made here is that the operation which he has described as an 'unwitting abstractive process' seems in most respects identical with the stage which we have called a perceptual analysis of the material. His term seems preferable and would be adopted by the writer.

It may be noted that the proposal to place less emphasis upon the constructive stage of recall, and more upon the original analysis which occurred with the first presentation, offers some partial explanation for the consistency of one reproduction with another under the conditions of the present experiment. At each stage of reproduction the important variable would not be the reconstructive process peculiar to that stage alone but the initial analysis common to all recall stages. The point would also have some bearing upon the importance of the initial reaction to the material in learning studies which we may now consider.

(v) *The importance of initial responses in learning*

It was noted in the section on Experimental Procedure that the time intervals were being manipulated to encourage slow learning in order that some estimate might be made of the relative strengths of particular variables. We will now see how this did in fact operate, and briefly consider the bearing of those results upon learning studies.

In passage A there was mentioned a specific sum of money, £100,000. It would be expected that errors would occur and they did. Some subjects could repeatedly reproduce an amount consistently in an erroneous variant; one subject gave six of his seven versions as £1000 and another gave five as £50,000. One subject did have a general selection of £20,000 (twice); then £10,000 and finally, £100,000. This last subject may represent the person who for a time just did not know what was the exact sum but was aware that it involved a big amount: but the examples of subjects repeatedly giving a wrong variant would seem to illustrate a different procedure. There are several possibilities which might account for this mistake:

(1) Subjects may have been so sure that they knew that particular part of the passage that they did not check it carefully when it was read out to them each week—the mistake persisted because they were not attending to the correct version.

(2) This is harder to conceive but perhaps subjects did observe the discrepancy between their version and the original but had forgotten it by the following week.

(3) A third possibility is that some observed the original without at first being aware of the discrepancy between it and what they had previously reproduced. Subsequently, it was their own version which they retained and reproduced in the following week. This would argue that the impression of their own version was stronger and more permanent than that of the original. On this third view learning would finally take place when a subject became aware of the discrepancy between the original and his own version.

On the evidence here it is not possible to be definite. But subjects who had repeated the same mistakes and had subsequently learned the correct version were questioned, and it was found that they did know both the correct answer and their previous error. Without claiming that the procedure of the third possibility was ubiquitous, it would be favoured as being at least prevalent. It would explain the occurrence of many of the mistakes on the same grounds as were found in the Serial Learning experiments (Kay, 1951), where complete learning was frequently taking place only after the identification of the error itself. In support of the generalization made there—learning involved unlearning—this experiment both illustrates it and extends its application. On the one hand, it applied to processes involving everyday speech habits—a word in the original was blocked by one from the subject's own vocabulary, e.g. 'urge' for 'impulse', 'purveyor' for 'seller', etc., and it was necessary for him to learn that his own word was wrong before he would replace it. On the other hand, there were broader, overall reactions to the material as a whole which a subject had made and perhaps because of their imprecision they were all the more difficult to modify; as noted in some cases subjects did not seem aware of their influence.

To conclude this section, the initial reproduction represented the interaction between what was given and what a subject brought to the task. The importance of this initial reaction was known from memory studies; it would seem that under these conditions where learning was possible its importance was no less. This was primarily because the subsequent modification which had to take place, if the material was to be learned, had to be undertaken by an organism which had already exhibited a predilection for the mistake which it had introduced. In learning it would be easy for one subject to amend the mistakes of another; it is altogether more difficult for a subject to amend his own. This experiment is in line with those which when considering the normal learning reactions of a mature organism have abandoned any S—R model and have placed the emphasis more and more upon what is already held, and less and less upon what is presented. What is initially laid down represents the interplay between 'permanent' directive tendencies within the organism and the presented material. Once that interaction has taken place the products are not easily modified. But there is a danger of overstating the case. In so far as learning generally demands complete accuracy, the difficulty of achieving it and the strength of what is already held is not misleading. But in noting the initial errors it is as well to be mindful of the general accuracy of the whole.

VIII. SUMMARY

An experiment was carried out with verbal material based on the method of repeated reproduction. (Weekly time-intervals.) But the standard procedure was modified in that the original version was read to subjects immediately after each reproduction.

Under these conditions it was found that:

1. The initial reproduction was a fairly accurate version of the content of the passage (70 %), but less so of the actual verbal material (30 %). The need for considering both criteria is brought out.

2. Subsequent versions were strikingly similar to each other and in their repetitions from the initial reproduction. They reached a stereotyped form comparable to that found in normal repeated reproduction experiments. This comparability applied to both content and verbal analysis.

3. The constructive nature of recall has been stressed by Bartlett but it is felt that in short-term studies emphasis must be placed upon the initial interpretation by a subject.

This interpretation is described as a 'perceptual analysis'. The process is compared to Gomulicki's recent study of an 'abstractive' operation. His term is preferred.

4. The relevance of the experiment to learning studies is brought out. Again it illustrates the necessity for an unlearning of mistakes, and in so far as the organism has shown a predilection for these erroneous introductions—either verbal or conceptual—the difficulty of their amendment is considerable.

APPENDIX A

Passage 'A'. Tea leaves

The actual tea leaves in the world are as varied and unstable as the actual political opinions of mankind. Every tea leaf in every tea garden is different from every other leaf, and a week of damp weather may change the whole stock in any warehouse. What therefore should the advertiser do to create a commercial 'entity', a 'tea' which men can think and feel about? He would choose some term, say 'Paramatta Tea', which would produce in most men a vague suggestion of the tropical East, combined with the subconscious memory of a geography lesson on Australia. He would then proceed to create in connexion with the word an automatic picture-image having previous emotional associations of its own. By the time that a hundred thousand pounds had been cleverly spent, no one in England would be able to see the word 'Paramatta' on a parcel without a vague impulse to buy, founded on a day-dream recollection of his grandmother, or of the British Fleet.

Passage 'B'. The stenographer

Wanted stenographic position—I take shorthand notes quicker than a new band of Oklahoma outlaws can spring up. I lay a smoke screen with a typewriter, and I do not stay long with one firm because I work so fast I burn the office down.

References: several large firms in Oklahoma City had complete new furniture of steel installed after I had worked there a few days, and the police raiding squad stationed me with a typewriter, and copying the encyclopaedia, in front of the door of a speakeasy to smoke out the bootleggers instead of throwing tear gas bombs into the building. Warning —I advise the person who contemplates hiring me to have his office heavily insured, because if I work for him he will have a sure case of conflagration on hand.

APPENDIX B

Content of the two passages

Passage 'A'. Tea leaves

Item 1. Tea leaves are as varied and unstable as political opinions.
2. Every tea leaf is different from another.
3. A week's damp weather changes a warehouse's stock.
4. How can an advertiser create a saleable object evoking thought and feeling?
5. Give the tea a name—Paramatta.
6. This name would suggest the Tropical East and memories of a geography lesson on Australia.
7. Create around the name a picture-image with an emotional association.

8. Spend £100,000 cleverly.
9. All England would want to buy when they saw the name 'Paramatta'.
10. This desire found on recollection of grandmother or British Fleet.

Passage 'B'. The stenographer

Item 1. Wanted stenographic position.
2. I take shorthand notes quickly (quicker than new band of Oklahoma outlaws can spring up).
3. I type very fast (so fast I lay a smoke screen).
4. I do not stay long with firms because I set the office on fire with working so fast.
5. References: Large firms in Oklahoma City.
6. Had steel furniture installed after I had worked there.
7. The police raiding squad put me outside a speakeasy.
8. The smoke from my typewriter was used instead of tear gas bombs on the bootleggers.
9. Warning: insure your office if you hire me.
10. I shall certainly cause you a fire.

There was no point in trying here to translate the content of passage 'B' into 'English'. But in marking the content of both reproductions no attention was paid to the literal accuracy of words.

REFERENCES

BARTLETT, SIR F. C. (1932). *Remembering.* Cambridge University Press.

GOMULICKI, B. R. (1952). Recall as an abstractive process. D.Phil. Thesis, University of Oxford.

KAY, H. (1951). Learning of a serial task by different age groups. *Quart. J. Exp. Psychol.* III, 166–83.

KAY, H. (1954). The effects of position in a display upon problem solving. *Quart. J. Exp. Psychol.* VI, 155–169.

(*Manuscript received 2 March* 1954)

CHANGES IN MOTOR RESPONSE WITH INCREASE AND DECREASE IN TASK DIFFICULTY IN NORMAL INDUSTRIAL AND PSYCHIATRIC PATIENT SUBJECTS[1]

By PETER H. VENABLES

Medical Research Council, Unit for Research in Occupational Adaptation, Maudsley Hospital

I. *Introduction* (p. 101). II. *Method* (pp. 101–104). III. *Results* (pp. 104–108). IV. *Discussion* (pp. 108–109). V. *Summary* (p. 109). *References* (p. 110).

I. Introduction

Clinical observation led Bleuler (1906) to describe two classes of people. In one, where 'sanguine' temperament is displayed, there is a quick and intensive reaction to emotional impression, in the other the affect is suppressed, leading to displacements and conversions. Nunberg (1918) produced experimental evidence for these two types in work on motor movement in response to word association. The first type made excessive, and the second decreased, motor movement under affect. Berrien (1939) found evidence of two extreme types of reaction to emotion-provoking stimuli in rhythmical finger movement of psychotic patients. He described these as excitatory and inhibitory in nature.

In work on the performance of pilots in a simulated aircraft cockpit, Davis (1948) observed two sorts of errors, those of overaction, and those of inertia. Subjects displaying overactivity reaction made larger control movements and excessive reaction to instrument deviation; the extent and gradient of movement was greatly increased and overcorrecting frequent. Subjects showing inertia reaction made errors which were large and of long duration; control movements were relatively small, and were less disturbed by restless movement than those in the performance of subjects showing overactivity reaction. As a further test of this finding of an 'overactivity-inertia' continuum, Davis devised a less complicated task involving pointer alignment. Similar reactions were observed on this task. The relationship of these reactions to personality type was found to confirm previous work. Anxious subjects were found to show overactive, and hysterics inert reactions.

The present experiment was designed to provide further data bearing on the hypothesis that (*a*) behaviour on a task may vary on a continuum ranging from overactive to inert; and (*b*) that this behaviour is related to personality factors in such a way that when the task is made more difficult dysthymics tend to show modification of performance in the overactive direction, while hysterics change to a more inert form of behaviour. In addition, further evidence was sought on the direction of change of behaviour when task difficulty is decreased.

II. Method

The experiment was conducted as part of a larger investigation into the effects of psychological handicap. It occupied approximately 8 minutes during a session lasting $1\frac{1}{2}$ hours during which the subject completed a battery of intelligence and personality tests. The short duration of the experiment was not deemed to be a disadvantage in that it offered a possibility to test previous workers' findings under conditions not previously employed, and provided a test of the possibility of generalization.

---

[1] Based on part of a thesis approved for the degree of Ph.D. in the University of London.

The type of experiment employed was dictated by the practical aim of simulating the stimulus conditions met with by a bus conductor in the course of his job. It provided a means of measuring the subject's reaction to an increase or decrease in the rate of presentation, or in complexity of stimuli. This was an attempt to parallel change in 'load' conditions imposed by an increase or decrease in the number of passengers boarding the bus.

Subjects. Sample 1 consisted of 210 male trainees for the job of bus conductor. Their age range was 18–46 years, mean $28{\cdot}1 \pm 7{\cdot}5$. Data were available for each subject to enable factor scores of Intelligence, Neuroticism, and Introversion-Extraversion to be allocated to him. These scores were derived from the analysis of results on 15 psychometric tests.

Sample 2 had the purpose of providing validation groups to substantiate findings obtained on sample 1. The subjects in sample 2 consisted of 11 showing anxiety states (age range 21–48 years, mean $30{\cdot}7 \pm 7{\cdot}3$) and 11 hysterics (age range 21–46 years, mean $32{\cdot}8 \pm 8{\cdot}5$) selected from a population of 52 male psychoneurotic patients interviewed at a psychiatric hospital. The selection was made by taking those subjects where clear psychiatric diagnoses showing absence of features other than anxiety or hysteria were confirmed by scores obtained on a questionnaire which had been found previously (Hildebrand, 1953) to differentiate anxiety states from hysterics. The two groups are known as Anxious (A) and Hysterics (H) for purposes of reference in the text. To enable results from samples 1 and 2 to be compared, two groups of 11 subjects each from sample 1 were selected on the basis of high scores on the neuroticism factor and extreme scores on the introversion-extraversion factor; these groups, chosen to have characteristics which tended to approach those of the anxious and hysteric hospital patient groups, may be called quasi-anxious (QA) and quasi-hysteric (QH). The terms quasi-hysteric and quasi-anxious have been adopted in view of the possibly misleading connotation of other labels which might have been used, e.g. industrial hysteric. The groups should be considered to be operationally defined, although it is hoped that they have some resemblance to those from sample 2.

Apparatus. From the subjects' viewpoint the apparatus consisted of a milk-white screen $8\frac{3}{4} \times 5\frac{1}{2}$ in., set in a brass frame, at an angle of 45° from the base. On either side of the centre line of the frame and at $2\frac{3}{8}$ in. from it were two marks over which the stimulus lights appeared. In the horizontal shelf in front of the display screen was a slot $8\frac{1}{2}$ in. long through which a knob protruded. A pointer attached to the knob indicated its position against the marks on the display panel. Five stimulus lights were set vertically behind the display panel above each lateral mark. Their position was not visible until alight. Reading from the bottom the stimuli were (1) white, (2) green, (3) red, (4) showing the word YES, (5) showing the word NO.

The subject was given the instructions, 'When you hear this click' (a signal indicating that new stimulus was being presented), 'move the pointer from the mark in the centre to the mark under the white light, if it is on its own, or has green, or YES over it. However, if it has red or NO over it move it to the mark on the opposite side. After this, move the pointer as quickly as possible to the centre mark'.

The pointer was attached to a lever controlling a pen which recorded the subject's movement on a moving paper. The stimuli were automatically presented in predetermined sequence via the contacts of three uniselectors. The presentation of stimuli was divided into three consecutive periods of 50 displays each. In the first easy period *A*, white lights only were used at a 2 sec. periodicity; this was followed by a difficult period *B*, using all the lights as 1·5 sec. per display, and a final easy period *C*, identical with *A*. The whole task thus took 275 sec. to complete. The order of presentation of lights was such that each direction of movement to each type of display was balanced. Every 5th response in the difficult period *B* was to a simple white light stimulus. This scheme was adopted as it was required to measure characteristics of response which were not themselves reactions to difficult stimuli, but were set in an environment of difficult stimuli. By this means it was hoped to measure not the immediate effect of distraction, but the effect on response to a standard stimulus of the emotional condition produced by having to respond to more complex stimuli.

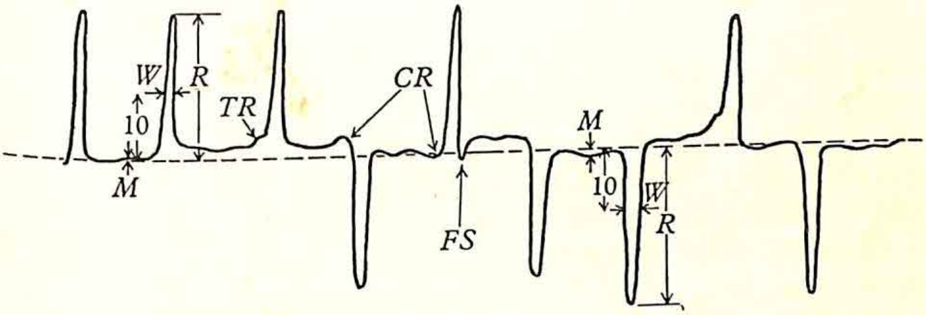


Fig. 1. Artificial example showing method of scoring psychomotor task records.

Procedure. The subject was seated in front of the apparatus which was placed on a low table, so that his forearm was parallel to the ground; he was asked to use his preferred hand. The task was explained to the subject, who was allowed to practise at his own speed whilst 50 displays were presented. It was then explained that his responses would be recorded and that he should expect a period of white lights only at a slow speed, a period using all the lights at a slightly faster speed, and finally white lights only again at a slow speed, there being no break between the periods.

The scoring of the records. Thirty of the 150 responses recorded for each subject were used for measurement. These were every 5th response and were to simple white light stimuli throughout. The method of marking is shown in Fig. 1. The following measures were used and are designated by symbols used throughout the remainder of the paper:

R	Maximum extent of response.
W	Width of response measured at a fixed distance (10 arbitrary units) from the centre line.
M	The deviation from the centre line of the point of return after each response.

Ten responses were measured in each period *A*, *B* and *C*, and means and standard deviations were calculated for each variable. The means were designated $\bar{R}$, $\bar{W}$ and $\bar{M}$ and the standard deviations, giving measures of variability within each period, were denoted by S_R, S_W and S_M.

In addition to this, measurement of selected responses secondary features of all the responses in each period were counted. These were:

FS False starts, where the subject started initially in wrong direction.

TR Trailing returns, where the subject did not make a simple direct return movement to the centre line after a response.

CR Corrected returns, where the return movement overshot the centre mark and had to be corrected.

III. RESULTS

(a) *Establishment of a method of measuring the overactive-inert continuum*

Prior to the commencement of any measurement, examination of records of the subjects in sample 1 showed that they ranged from those showing extensive responses of short duration with considerable overshooting and subsequent correction, to those showing short responses lasting some length of time. These were thought to represent opposite poles of the overactive-inert continuum. A period of 25 typical responses (nos. 10–34) in period *A* was rated for this feature on a five-point scale by the writer and an assistant. The correlation between the ratings was 0·807. There was thus a feature of the records which could be determined with reasonable reliability. The problem was to decide how this feature might best be measured so that changes between the different periods of the task could be followed. The possible measures of activity could be derived from those measures already outlined; these were:

(1) The amount of response per unit time, measured by $\bar{R}/\overline{W}$.

(2) The magnitude of the response, measured by $\bar{R}\times\overline{W}$.

The term magnitude of response is somewhat ambiguous and needs further explanation. Measured by $\bar{R}\times\overline{W}$, it is proportional to the area of the response. Thus a response of small magnitude is one of small extent, or one which is completed quickly, or both. The other measure $\bar{R}/\overline{W}$, where the extent of response is divided by the time taken for it, gives a measure which is of course related to 'magnitude' but is not identical. A response showing 'large response per unit time' is one where a response of large extent is completed quickly. We should expect a positive correlation between small magnitude of response and large response per unit time, although in so far as there is a difficulty of choice between the two measures this correlation should depart from unity. The problem of deciding which of these measures was more suitable was tackled in the following way.

It was thought that use could be made of the secondary features of responses along with the rating determined above to provide evidence on which a decision might be made. It was hypothesized that an 'active' subject might show a greater number of false starts (*FS*) and corrected returns (*CR*), and a smaller number of trailing returns (*TR*) than an inert subject. With this in mind $\bar{R}/\overline{W}$, $\bar{R}\times\overline{W}$, *FS*, *TR*, *CR*, and the rating of overactivity-inertia (Rat. O-I), scores for period *A* were intercorrelated. The table of intercorrelations (Table 1) shows that R/W has the highest correlation with the rating, and in addition that the pattern of correlations of R/W with *FS*, *TR* and *CR* follows more closely that of the rating with these variables than does that of $R\times W$. These factors determined the choice of R/W as a suitable measure of 'activity'. To add further confirmation a Lawley (1943) maximum likelihood type factor analysis was carried out on the intercorrelation between R/W_A, FS_A, TR_A, CR_A and the rating. The result is shown in Table 2.

Table 1. *Intercorrelation of hypothesized measures of overactivity-inertia, sample* 1, $N=210$

		R/W_A	$R \times W_A$	FS_A	RT_A	CR_A	*Rat. O-I*
Large response per unit time	R/W_A	—	—	—	—	—	—
Small magnitude of response	$R \times W_A$	0·694	—	—	—	—	—
Many false starts	FS_A	0·081	0·500	—	—	—	—
Few trailing returns	TR_A	0·221	−0·078	0·337	—	—	—
Many corrected returns	CR_A	0·491	0·237	0·040	0·129	—	—
Overactivity	*Rat. O-I*	0·858	0·618	0·001	0·379	0·437	—

Note: A correlation of 0·136 is significant at the 0·05 level.

Table 2. *Unrotated factor loadings of hypothesized measures of overactivity-inertia*

		I°	II°
Large response per unit time	R/W_A	0·924	0·269
Many false starts	FS_A	0·089	−0·226
Few trailing returns	TR_A	0·342	−·0231
Many corrected returns	CR_A	0·511	0·023
Overactivity	*Rat. O-I*	0·905	0·184
	% variance accounted for	41·18	4·4

The first unrotated factor accounts for 41·18% of the variance, and is most highly loaded on R/W and the rating. It was not possible to identify the second factor, nor was it thought worth while to extract further factors which were shown to be present in the residual matrix by Lawley's test of significance.

The variable R/W_A resulted from measures taken on responses 1, 6, 11, 16, etc. A measure of reliability can be calculated by comparing this score with a similar one taken from responses 3, 8, 13, ..., etc. When this was done a reliability coefficient of 0·970 was obtained. It is clear from this work that a factor of overactivity-inertia was present in the records and that it could be reliably measured.

It was hoped that a measure of 'disorganization of behaviour' independent of the 'activity' factor described above might be developed. The hypothesis could be advanced that a record showing organization was one where the responses, however large or small, were of even length and width and showing accuracy of alignment to the centre line. Disorganization would then be characterized by high M, S_R, S_W and S_M scores. Following the previous procedure the records were rated on a 5-point scale for 'disorganization of response'. The reliability of the rating was 0·667. Measures S_R, S_W, S_M, M and the rating were intercorrelated, with R/W as a marker for the activity factor. The factor analysis of the matrix yielded a first factor accounting for 23·5% of the variance. This having its highest loading on R/W was again clearly an activity factor, the second factor could not be interpreted in an orthogonal position, and when an oblique solution through the cluster S_R, S_M, and the rating was adopted, the correlation between the factors was approximately 0·87. It was therefore apparent that no measure of disorganization could be developed independently of the 'Activity' factor.

(*b*) *Determination of the relation of the overactive-inertia continuum of motor behaviour to personality classification.*

Five types of measures were calculated for each subject for each of the nine variables outlined earlier and for the index R/W. These were:

(1) Performance in period A, e.g. R/W_A.
(2) Performance in period B, e.g. R/W_B.
(3) Performance in period C, e.g. R/W_C.
(4) Change in performance between periods A and B, e.g. R/W_{A-B}.
(5) Change in performance between periods B and C, e.g. R/W_{B-C}.

These measures were correlated with factor scores measuring Intelligence, Neuroticism, and Introversion-Extraversion for the group of 210 normal subjects, sample 1.

It was found that increase in Inertia from period A to period B measured by R/W_{A-B} was positively correlated with the factor of Extraversion at a significant level ($r = +0{\cdot}141$, $P < 0{\cdot}05$), and that increase in inertia between period B and period C (R/W_{B-C}) was negatively correlated with Extraversion ($r = -0{\cdot}147$, $P < 0{\cdot}05$). No significant correlations with the other two personality factors were observed.

In the case of the neurotic group, sample 2, rank correlations between the same variables and the anxious-hysteric dichotomy were calculated. It was found that inertia in period A (R/W_A) was significantly correlated with hysteria ($\tau = +0{\cdot}36$, $P < 0{\cdot}05$) and that increase in inertia from period B to C was negatively correlated with hysteria ($\tau = -0{\cdot}41$, $P < 0{\cdot}05$).

The results from 22 subjects from sample 1 selected to be comparable with the anxious and hysteric groups showed that, as in the parent population, increase in inertia from periods A to B, was correlated with hysteria (neurotic–extraversion) ($\tau = +0{\cdot}57$, $P < 0{\cdot}05$) and increase in inertia from periods B to C was negatively correlated with hysteria, ($\tau = -0{\cdot}63$, $P < 0{\cdot}05$).

The rank correlation method used is that developed by Whitfield (1947) from Kendall's (1938) work, for use where one variable is ranked and the other dichotomous. It should be interpreted in this case in the same manner as a biserial coefficient. The use of this method was dictated by the non-normal distribution resulting from the examination of extreme cases.

Mean scores for R/W, R, and W for periods A, B and C, and for change in performance between periods A and B, and B and C are given in Table 3. These are shown for the total normal sample for groups QA and QH derived from it, and for groups A and H from the psychiatric patient sample, sample 2. Mean scores which are significantly different in groups QA and QH and groups A and H in italics. The significance of the difference between scores on these sub-groups was tested by the median method (Moses, 1952); in view of the non-normal distribution of measures. Fig. 2 shows in graphic form the relationships outlined above and shown in the upper part of Table 3.

If we examine the results from either the whole of sample 1 or the subgroups drawn from it, it is seen that extravert and quasi-hysteric subjects show a significant tendency to change their performance in the inert direction under increased difficulty. The reverse tendency is shown in introvert, quasi-anxious, and anxious subjects The hospitalized, hysteric patient presents an initially inert type of response. Commencing performance on this level he does not change his performance in the inert direction under increased difficulty, but maintains a performance which is substantially similar to his starting performance.

Table 3. *Mean performance scores of subjects from samples 1 and 2 on variables R/W, R, and W, for periods A, B, and C*

	Sample 1			Sample 2	
Variable	Total	QA	QH	A	H
R/W_A	5·93	5·32	6·05	*6·09*	*4·33*
R/W_B	5·38	5·71	5·59	7·03	4·39
R/W_C	5·58	4·82	6·00	6·08	4·72
R/W_{A-B}	+0·55	*−0·39*	*+0·46*	−0·94	−0·06
R/W_{B-C}	−0·20	*+0·89*	*−0·41*	*+0·95*	*−0·33*
R_A	21·37	20·97	21·68	*20·66*	*19·35*
R_B	20·62	20·50	20·73	19·84	19·10
R_C	21·40	21·38	21·78	20·55	19·76
R_{A-B}	+0·75	+0·47	+0·95	+0·82	+0·25
R_{B-C}	−0·78	−0·88	−1·05	−0·71	−0·66
W_A	3·93	4·33	3·76	4·30	4·95
W_B	4·18	3·83	3·92	*3·76*	*4·83*
W_C	4·17	4·66	3·82	4·56	4·57
W_{A-B}	−0·25	*+0·50*	*−0·16*	+0·54	+0·12
W_{B-C}	+0·01	*−0·83*	*−0·10*	*−0·80*	*+0·26*
N	210	11	11	11	11

Differences significant at less than the 0·05 level of confidence are in italic.

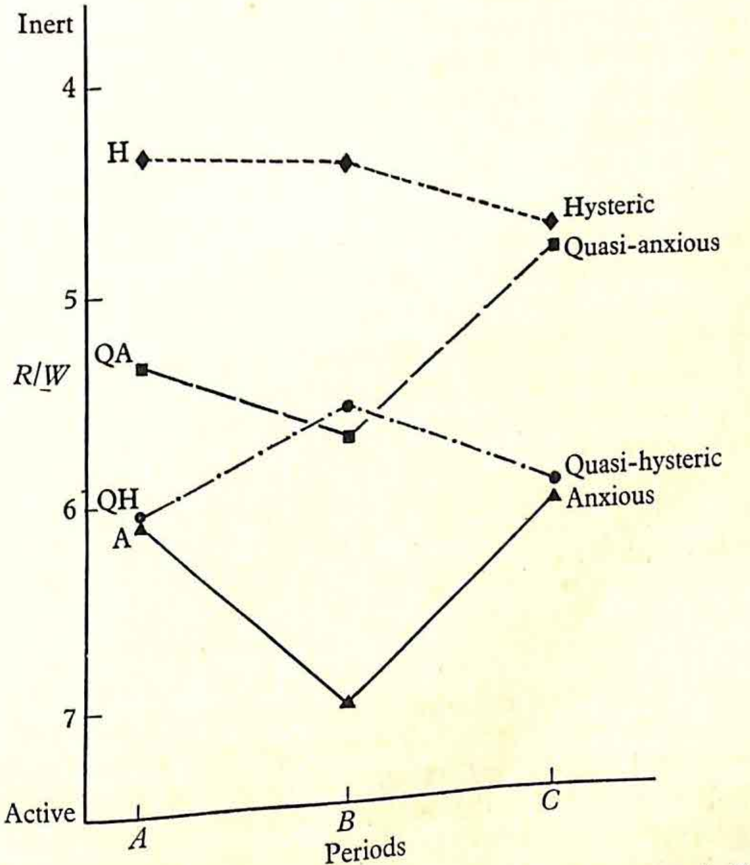


Fig. 2. Changes in activity of response (R/W) between periods A, B, and C in groups A, H, QA and QH.

With decrease in task difficulty there is in all cases a significant tendency for the extravert or hysteric subject to perform more actively whilst the introverted or anxious patient subject shows a decrease in activity.

On examination of the measures which make up the index R/W it is seen that response length tends to shorten in the difficult period, and lengthen subsequently, although to different extents, in different groups of subjects. On decrease of difficulty there is a general lengthening of response in all groups. It is the measure W, the time for which the response lasts, that seems to be responsible to a greater extent for differential behaviour between the groups. In both dysthymic groups there is a speed up of response under difficulty followed by a slowing down with decrease in difficulty. In the hysteric patient group performance is slow initially whilst in the extraverted group drawn from sample 1 there is a slowing down on increase of difficulty. In both of these groups there is a speed up on decrease of difficulty.

While W, measuring speed of response, appears to parallel many of the findings of R/W in dysthymic-hysteric group differentiation it is felt that the latter measure does operate sufficiently differently to justify the reasons for its original inclusion. Evidence is given when motor task variables are related to G.S.R. measures. The work of White (1930) Freeman & Simpson (1938) and Staudt & Kubis (1948) shows that increase in G.S.R. conductance accompanies increase in muscular tension. Other work reported elsewhere (Venables, 1955), indicates that increase in 'activity' measured by R/W_{A-B} is significantly correlated ($P=0{\cdot}045$ and $0{\cdot}03$) with increase in conductance, in both normal and neurotic groups. Increase in speed measured by W_{A-B} does not show significant correlations with increase in conductance. The work of Freeman & Pathman (1942) and Champion (1950) would lead us to expect that an increase in motor movement would result in a dissipation of autonomic disturbance. In the case of G.S.R. response after the difficult period, it is the measure W_{B-C} and not R/W_{B-C} which shows a significant correlation with decrease in conductance supporting Freeman & Pathman's and Champion's work. It would thus seem that there is a useful differentiation between the measures R/W and W, and that the former may measure behaviourally something akin to muscular tension, whilst the latter measures speed of response.

Correlations of task variables with Intelligence and Neuroticism show that they are in general independent of these factors. If a group of subjects from sample 1 having ambivert and low neuroticism factor scores are compared with the two patient groups the finding of independence of the overactivity-inert continuum from neuroticism is supported by insignificant rank correlations of $+0{\cdot}05$, $+0{\cdot}18$ and $0{\cdot}00$ with R/W_A, R/W_{A-B}, and R/W_{B-C} respectively. Independence from age is shown by correlations of $+0{\cdot}09$, $+0{\cdot}01$ and $+0{\cdot}01$ with the same variables.

IV. DISCUSSION

In the work described an attempt has been made to provide further data bearing on the relation between motor performance and personality. The work of previous writers has been largely confirmed, and additional findings showing evidence of regulatory behaviour after task difficulty has ceased have been given. It is of some interest that in many ways it is the change in performance when the stimulus situation changes from difficult to easy, which seems to be the most valuable indicator of different functions in different types of personality. Whilst in the hysteric group there was little change when the task difficulty increased, it was the marked change in the direction of overactivity on decrease of difficulty which differentiated this group from the anxious patients.

An external source of evidence on this point is given when the performance scores of 18 subjects in sample 1 who were discharged (largely for dishonesty) are examined. The scores for these subjects are compared with those resigning or remaining in employment in the same period of six months. It was found that the measure R/W_{B-C} differentiated the discharged from the rest at a significant level ($P<0{\cdot}01$). This was not, however, the case with R/W_{A-B} where a quite insignificant F-ratio was obtained. The discharges show behaviour which was in many ways similar to that of extraverted subjects.

The results of this work have some similarities to those of Foulds (1952) although he used a very different type of experiment. Foulds found that distraction produced a speed up of performance of dysthymics on the Porteous Maze, whilst speed up, although shown in non-dysthymics, was not significant. Examination of Table 3 shows a very similar state of affairs in the performance of our anxious and hysteric groups on periods A and B.

Also shown in the data presented is some indication of the difference between the performance of hospitalized and non-hospitalized groups of tentatively similar personality type. This is particularly striking in the case of the hysterics and quasi-hysterics, where the initial performance of the hysteric in hospital is so inert that no change in the inert direction seems possible. On the other hand, the quasi-hysteric in industry presents an initially quite active performance. These differences should not be pursued too far in view of the present unsatisfactoriness of tests with which it is hoped to measure in a normal population the counterpart of anxiety-hysteria in a hospital population.

The findings quoted indicate that performance is by no means a static phenomenon and that perhaps greater attention should be given in industrial situations to the maintenance of an even level of difficulty of work, in order to avoid alteration of performance due to interaction effects with features of the operatives' 'personality' in its narrowest sense.

Any attempt to present a theoretical basis for these findings is at the moment considered premature; further experimentation to test various hypotheses is proceeding.

V. SUMMARY

A motor task, involving the movement of a pointer in relation to stimulus lights under easy, difficult and subsequently easy conditions was performed by 210 male subjects consisting of a random selection of trainee bus conductors, age range 18–46, and by 22 male patients, age range 21–48, in a psychiatric hospital, selected from a group of 52 patients as having clear-cut psychiatric and questionnaire diagnosis.

Performance on the motor task could be described as falling on a continuum of overactivity-inertia. In general, under conditions of increased task difficulty, extraverted subjects showed a tendency to increased inertia of performance and introverted subjects to increased activity. When the task became easier the reverse relationship was found to hold true.

The author wishes to acknowledge the valuable assistance received from Dr H. J. Eysenck and Mr M. B. Shapiro and his colleagues Drs Eisler, Heron, Markowe, O'Connor and Tizard. The psychiatric patient subjects were made available through the co-operation of Dr M. Desai.

REFERENCES

BERRIEN, F. K. (1939). Finger oscillations as indices of emotion. I. Preliminary validation. *J. Exp. Psychol.* XXIV, 485–98.

BLEULER, E. (1906). *Affektivität, Suggestibilität, Paranoia.* Halle: Verlag Carl Marhold.

CHAMPION, R. A. (1950). Studies of experimentally induced disturbance. *Aust. J. Psychol.* II, 90–9.

DAVIS, D. R. (1948). *Pilot Error—Some Laboratory Experiments.* London: H.M.S.O.

FREEMAN, G. L. & PATHMAN, J. H. (1942). The relation of overt muscular discharge to physiological recovery from experimentally induced displacement. *J. Exp. Psychol.* XXX, 161–74.

FREEMAN, G. L. & SIMPSON, R. M. (1938). The effects of experimentally induced muscular tension upon palmar skin resistance. *J. Gen. Psychol.* XVIII, 319–26.

FOULDS, G. A. (1952). Temperamental differences in maze performance. Part II. The effect of distraction and of electroconvulsive therapy on psychomotor retardation. *Brit. J. Psychol.* XLIII, 33–41.

HILDEBRAND, H. P. (1953). A factorial study of introversion-extraversion by means of objective tests. Ph.D. Thesis, University of London.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, XXX, 81–93.

LAWLEY, D. N. (1943). The application of the Maximum Likelihood Method of Factor Analysis. *Brit. J. Psychol.* XXXIII, 172–5.

MOSES, L. E. (1952). Non-parametric statistics for psychological research. *Psychol. Bull.* XLIX, 122–43.

NUNBERG, H. (1918). On the physical accompaniments of association processes. In Jung, C. G., *Studies in Word Association.* London: Heinemann.

STAUDT, V. & KUBIS, J. F. (1948). The psychogalvanic response and its relation to changes in tension and relaxation. *J. Psychol.* XXV, 443–53.

VENABLES, P. H. (1955). Some findings on the relationship between G.S.R. and motor task variables (To appear in *J. Gen. Psychol.*).

WHITE, M. M. (1930). Relation of bodily tension to electrical resistance. *J. Exp. Psychol.* XIII, 267–77.

WHITFIELD, J. W. (1947). Rank correlation between two variables one of which is ranked, the other dichotomous. *Biometrika*, XXXIV, 292–6.

(*Manuscript received* 21 *April* 1954)

THE IDEA OF A DISTRIBUTION

By JOHN COHEN and C. E. M. HANSEL

Department of Psychology, University of Manchester

I. *Introduction* (p. 111). II. *Method* (pp. 111–112). III. *Type I: procedure and results* (pp. 112–115). IV. *Type II: procedure and results* (p. 116). V. *Type III: procedure and results* (pp. 116–120). VI. *Discussion of results* (pp. 120–121).

I. INTRODUCTION

The experiments to be described form part of the study of what we propose to call 'subjective probability'. In everyday life we are constantly choosing, estimating, predicting, judging, or deciding on the basis of incomplete knowledge. In all these activities there is an element of uncertainty on the part of the individual. In so far as the individual prefers one alternative to the other or considers it more likely to a greater or less extent, his preferences may perhaps be evaluated. We shall inquire whether judgements based on partial information follow their own psychological rules in more or less consistent fashion, and how these rules compare with rules of mathematical probability based on logical criteria. A second aim of this inquiry is to trace the systematic or other changes in the forms of subjective probability characteristic of different ages during the course of development. It is of particular interest to examine the way the child assimilates statistical rules in so far as they are repeatedly verified in experience from the earliest days of life. In this first series of experiments we shall confine ourselves to the age range 10 to 15 years. We should add at this point that we are here concerned with the judgement as such. In further elaboration of these studies it will be important to distinguish between the judgement itself and the intensity of conviction with which it is held.

II. METHOD

In the present paper we examine the types of 'guess' which are made in situations of increasing complexity. These range from choices between two alternatives to judgements which may implicitly involve ideas of a frequency distribution. We also refer to conditions in which the amount of information provided is systematically controlled. In one experiment, for example, we show a group of subjects a large urn containing an unspecified number of blue or yellow beads. They are not told the proportions of blue and yellow beads in the urn. The beads are thoroughly mixed and we take out 4 and place them in a beaker. This is repeated until there are 16 beakers each containing 4 beads. The subjects might then be asked how many of the beakers contain, say, exactly 2 blue beads. From the data given by the subjects, the distribution of beakers containing 0, 1, 2, 3, and 4 blue (or yellow) beads can be built up and compared with a theoretical distribution.

The particular distributions to be studied in these experiments may be built up from subjects' estimates in three different ways. These vary with the kind of judgement made by the subject. The three different ways are:

Type I. The beads are withdrawn from the urn in sets of 4. They are placed individually in separate beakers and the subject guesses the colour of each bead.

Type II. The beads are withdrawn from the urn in sets of 4 and each set of 4 is placed in a separate beaker. There are 16 beakers in all. The subject guesses the number of blue and yellow beads in each beaker.

Type III. The beads are placed in 16 separate beakers, as in II above. The subject now guesses the number of *beakers* containing 0, 1, 2, 3 and 4 beads respectively.

In the present series of experiments, which are analysed in detail, the proportions of blue and yellow beads in the urn are stated to the subjects to be unknown. We refer briefly to other experiments in which the subjects are told that definite proportions of beads are in the urn.

The total number of subjects studied in these experiments is about 330. About 100 subjects took part in experiments relating to Types I and II and about 230 in experiments relating to Type III.

III. TYPE I: PROCEDURE AND RESULTS

The procedure takes the form of a class demonstration in which beads are drawn singly from a large urn containing blue and yellow beads in unspecified numbers, and the subjects guess the colour of each bead.

We say to the subjects:

> There are a lot of beads in this urn but we do not know exactly how many. They may be blue or yellow, but no other colour. There may be no blue beads or any number of blue beads. There may be no yellow beads or any number of yellow beads. We do not know the proportions of yellow and blue beads in the urn. We shall take 4 beads at a time from the urn and place each one of these in a separate beaker. There are 4 beakers here, one for each bead.

The subjects can hear the beads drop into the beakers but they cannot see the colour of the beads at any stage of the experiment. They record their guess of the colour of the bead in each of the four beakers. The beads are then replaced in the urn. This procedure is repeated 16 times.

The 95 subjects yield 1520 arrangements of a set of 4 beads. In Table 1 we show in col. (i) the 16 possible arrangements of the 4 beads. Col. (ii) gives the number of times each of the 16 arrangements occurs in the choices made by the group. Col. (iii) gives the theoretical number of times each arrangement would occur if the ratio of blue to yellow beads were based on the respective number allotted by the group (3162 blue and 2918 yellow) and not on a 1:1 ratio. On the assumption of a 1:1 ratio the theoretical number of times each arrangement would occur would invariably be 95. The values in col. (iii) may not be directly comparable with those in col. (ii) because individual subjects may be basing their choices on implicit assumptions about differing proportions of blue and yellow beads in the urn. In so far as such implicit assumptions are made they may fluctuate from time to time during the course of the experiment. The effect of these varying assumptions would be to give a flattened theoretical distribution as compared with the distribution based on group values. In view of these considerations we show in col. (iv) the theoretical number of times each of the 16 arrangements will occur if the ratio of blue to yellow beads is based on the number of each colour allocated by each individual rather than on the totals of the group as a whole. The values in col. (iv) were obtained by calculating a binomial distribution for each individual in terms of the proportions of blue and yellow beads derived from his total choices, then summing the results for all 95 subjects.

To return for a moment to col. (iii) in Table 1, we should note that the proportions of blue and yellow beads in the urn were stated by us to be unknown. We find that the proportions allotted by the subjects tended on the whole towards 0·5, the actual values for the group being 0·52 (blue) and 0·48 (yellow) with a standard deviation of 0·06. We should expect that unless any special influence affected the number of choices of one rather than the other colour, the proportions implicitly allotted to them would tend to 0·5. In fact the proportions 0·52 and 0·48 depart significantly from this mean value. From the observed number of choices, 3162 blue and 2918 yellow, χ^2, calculated from the expected number, 3040, is 9·8 ($P<0{\cdot}01$). This may be due to a number of causes such as colour preference, a preference for writing *B* rather than *Y*, or to the wording of the instructions.

Table 1. *Choice of all possible arrangements of 4 beads, each of which may be blue or yellow*

(N = 95. Mean age 10 years 6 months.)

	Frequencies		
Arrangements (i)	Observed (ii)	Calculated from proportions of colours allotted by the group* (iii)	Calculated from proportions of colours allotted by each individual (iv)
B B B B	151	111·1	123·8
B B B Y	116	102·6	102·9
B B Y B	83	102·6	102·9
B Y B B	77	102·6	102·9
Y B B B	99	102·6	102·9
B B Y Y	102	94·7	92·0
B Y B Y	93	94·7	92·0
B Y Y B	88	94·7	92·0
Y B B Y	123	94·7	92·0
Y B Y B	75	94·7	92·0
Y Y B B	76	94·7	92·0
B Y Y Y	74	87·4	86·7
Y B Y Y	62	87·4	86·7
Y Y B Y	90	87·4	86·7
Y Y Y B	93	87·4	86·7
Y Y Y Y	118	80·7	85·6

* These expected values are calculated from a population of blue and yellow beads in the proportions 0·52 and 0·48 respectively, representing the colour preferences of the group.

If we now compare the observed values in col. (ii) with those shown in col. (iv), calculated on the basis of proportions of colours allotted by each individual, we find that they differ significantly. χ^2 is 57·25, D.F. = 15, $P<0{\cdot}001$. This means that even when we take the individual allotments of colour fully into account the distribution of arrangements is not random. Other factors must therefore be operative. Inspection of the values in col. (ii) suggests that one such factor may be a preference for particular kinds of arrangement.

We can examine this suggestion in the first instance by comparing the preference for different arrangements within each of the five subgroups, each subgroup containing 0, 1, 2, 3 or 4 blue beads, disregarding order. Within any subgroup, if there were no preference for any particular arrangement, the expected number for each arrangement would be the same. We calculate the expected values from the total number of choices within the subgroup divided by the number of possible arrangements in the particular subgroup (see Table 2). $\chi^2=35{\cdot}1$, D.F. = 15, $P<0{\cdot}01$, thus indicating a preference for particular arrangements.

The data indicate thus far that choices are not made at random but vary in a systematic manner within each subgroup where the number of beads of each colour remains constant. We may now ask what it is that determines this preference for particular kinds of arrangement.

Table 2. *Choice of arrangements of* 0, 1, 2, 3 *or* 4 *blue beads disregarding order*

(N = 95 Mean age 10 years 6 months.)

Arrangements (i)	Frequencies: Observed (ii)		Frequencies: Calculated by equal division of subgroup total (iii)
B B B B	151	151	151·00
B B B Y	116	375	93·75
B B Y B	83		93·75
B Y B B	77		93·75
Y B B B	99		93·75
B B Y Y	102	557	92·83
B Y B Y	93		92·83
B Y Y B	88		92·83
Y B B Y	123		92·83
Y B Y B	75		92·83
Y Y B B	76		92·83
B Y Y Y	74	319	79·75
Y B Y Y	62		79·75
Y Y B Y	90		79·75
Y Y Y B	93		79·75
Y Y Y Y	118	118	118·00
Total	1520	1520	1520·00

One possible factor affecting the preference for certain arrangements is their balanced or symmetrical character. From the point of view of symmetry we distinguish three kinds of arrangement:

(i)	B B B B	or	Y Y Y Y	balanced
	B Y Y B	or	Y B B Y	
(ii)	B Y B Y	or	Y B Y B	alternating or repetitive
	B B Y Y	or	Y Y B B	
(iii)	B B B Y	or	Y Y Y B	asymmetrical
	B B Y B	or	Y Y B Y	
	B Y B B	or	Y B Y Y	
	Y B B B	or	B Y Y Y	

When the observed values for the three groups are separately compared with the corresponding frequencies in col. (iv) of Table 1, we find that the observed value of the first group exceeds its theoretical ($\chi^2=19$, D.F. $=1$, $P<0\cdot001$); the observed value of the second is not appreciably different from its theoretical; the observed value of the third group is significantly lower than the theoretical ($\chi^2=5\cdot46$, D.F. $=1$, $P<0\cdot02$). This seems to indicate a preference for the balanced and an avoidance of the asymmetrical kind of arrangement.

Another factor might be the tendency in any given arrangement to repeat the previous choice of blue or yellow rather than to change to the other colour. The total number of choices, 1520, yields 4560 possible changes, i.e. three for every arrangement. There are in fact 2110 choices in which there is a change of colour and 2450 choices in which the colour is not changed. This difference is significant ($\chi^2 = 27{\cdot}1$, D.F. $= 1$, $P < 0{\cdot}0001$) and indicates a preference for choosing the same colour after any given choice within a set of 4 beads.

Although the subjects make their choices in relation to individual beads in sets of 4, we can combine the sets in such a way as to yield a distribution of arrangements containing 0, 1, 2, 3 or 4 blue beads. The subjects' choices classified in this way are shown in Table 3, together with the corresponding values calculated on the basis of the proportions of beads of each colour allotted by each subject.

Table 3. *Distribution of arrangements containing* 0, 1, 2, 3 *or* 4 *blue beads*

	Number of blue beads					
	0	1	2	3	4	Total
Observed values						
Total	118	319	557	375	151	1520
Mean	1·2	3·4	5·9	3·9	1·6	16
Calculated values*						
Total	85·6	346·8	552·0	411·6	123·8	1520
Mean	0·9	3·7	5·8	4·3	1·3	16

* Based on proportions of blue and yellow allocated by each subject.

If the subjects were making their choices at random, even if their preferences for the two colours were uneven, the two distributions in Table 3 would approximate to each other. The fact that the distributions differ significantly implies that the subjects are not behaving in random fashion. We have already suggested two factors that might account for this.

The comparison between the two distributions in Table 3 is given not to suggest that the children in this experiment had any notion of a binomial distribution, but to demonstrate that they were not merely allocating colours at random. Any resemblance between the two distributions may be due to the fact that apart from the factors isolated there is a very large and diverse number of influences affecting the choices of different individuals about which we know little. The resulting distributions accordingly tend in some measure to have the form of a distribution resulting from random events.

In the foregoing experiment the children guessed the colour of single beads. We might have expected that their choices would have been random and hence the distributions derived from the choices would tend towards the binomial. This does not appear to be the case. On the contrary we find:

(i) The proportions of blue and yellow beads implicit in the choices although varying to some extent have a relatively small scatter. Two-thirds of the proportions (for blue) are within the range 0·46 to 0·58.

(ii) Their choices are not made at random.

(iii) The non-random character of the individual choices is due to preferences for certain types of arrangement within the sets of 4 beads.

(iv) There is a preference for balanced and an avoidance of asymmetrical arrangements.

(v) There is a preference for repeating the choice of any one of two colours within any set of choices.

IV. TYPE II: PROCEDURE AND RESULTS

We now turn to the second way in which the distributions may be built up. Four beads at a time are placed in turn in each of the 16 beakers and, as they are heard to drop in the beaker, the subjects record their answer to the question (Exp. II*a*): how many blue beads are in the beaker? In Exp. II*b* the subjects were asked to record how many blue *and* yellow (or, reversing the order of the question, yellow *and* blue) beads they thought were in each beaker. The order II*a*, II*b* was followed with one group of subjects and II*b*, II*a* with a second, and the results for each experiment combined.

The distribution of arrangements yielded by the choices is shown in Table 4, together with the calculated values based on proportions of blue and yellow implicitly allotted by each individual.

Table 4. *Observed and calculated distributions*

(N = 93. Age 10 + years.)

Experiment		Number of blue beads					
		0	1	2	3	4	Total
II*a*	Observed	92	334	436	363	263	1488
	Calculated	65	283	518	459	163	1488
II*b*	Observed	245	318	361	310	254	1488
	Calculated	117	366	513	365	125	1488

Unlike the Type I experiment the subjects in Type II actually state a distribution for each one of the 16 beakers. Random behaviour on their part would yield an equal number of beakers for 0, 1, 2, 3 and 4 beads. This does not appear, for both Exps. II*a* and II*b* differ significantly from a rectangular distribution. Indeed they show a marked tendency in the direction of a binomial distribution. The subjects' choices imply that they regard 2B and 2Y as more likely than 1 B and 3 Y (or vice versa) or 0 B and 4 Y (and vice versa). Similarly, they regard 1 B and 3 Y or 1 Y and 3 B as more likely than 0 B and 4 Y (or 0 Y and 4 B). In this experiment random choices would lead away from a binomial towards a rectangular distribution. In the previous experiment random behaviour would lead away from a rectangular towards a binomial distribution of arrangements.

One other feature of this experiment should be noted, i.e. the difference between the distributions yielded by Exps. II*a* and II*b* respectively. This may be attributed to the form of the question. It appears that when we ask: 'How many blue beads are in each beaker?' and do not make any reference to yellow beads, the subjects tend to avoid the value 0 blue beads.

We may conclude from this experiment that subjects do not allocate proportions of blue and yellow beads within sets of 4 at random. They exhibit a significant preference for the value 2 B, 2 Y and also for 1 B, 3 Y or 3 B, 1 Y as compared with 0 B, 4 Y or 0 Y, 4 B.

V. TYPE III: PROCEDURE AND RESULTS

The third way in which a distribution may be built up is by placing 4 beads drawn at random into each one of 16 beakers. In this experiment the subjects are told that the proportions of blue and yellow beads in the urn are unknown. They now guess the number of beakers containing 0, 1, 2, 3 or 4 blue (or yellow) beads, and these guesses constitute their distributions. In one group (aged 11–13) the subjects, after giving their guesses, explained in writing why they chose their particular values.

In previous studies we have found that the order in which the questions are asked in this type of experiment may affect the subjects' estimates, the earlier questions tending to receive higher values than they would otherwise get. The estimates may also be affected by the order in which the colours are referred to in the questions. In order to eliminate these influences it was necessary to employ ten ways in which the questions were presented. In five of these blue was referred to first in the questions and in five yellow was referred to first. In the set of five presentations for each colour the questions were so arranged that all five questions appeared in each position an equal number of times. The ten orders of presentation were given to approximately equal groups of subjects, each subject receiving a single order of presentation. The means for the entire samples were obtained from the subgroup means.

A specimen set of questions is set out below:

NAME: .... Age: ....
How many beakers will have 4 yellow beads and no blue beads? ....
How many beakers will have no yellow beads and 4 blue beads? ....
How many beakers will have 3 yellow beads and 1 blue bead? ....
How many beakers will have 1 yellow bead and 3 blue beads? ....
How many beakers will have 2 yellow beads and 2 blue beads? ....
Total 16

The mean distributions for the different age groups are given in Table 5.

Table 5. *Beakers containing* 0, 1, 2, 3 *or* 4 *blue beads*

(Mean values allotted as proportions of 16. Unknown proportions of blue and yellow beads in the urn.)

		Number of blue beads					Total number of beakers
Age	N	0	1	2	3	4	
10 +	87*	3·2	3·1	3·7	2·9	3·2	16
12 +	31†	3·3	2·7	3·5	3·2	3·1	16
14 +	26†	3·0	3·0	3·7	3·2	3·2	16

* Including some 11-year-olds.

† Girls only.

We wish to see first whether the subjects, when the proportions are unknown, estimate the number of beakers containing 0, 1, 2, 3 or 4 blue beads respectively in a manner that is likely to arise in practice on any assumption that might be made about the contents of the urn. If, for example, a subject estimated that each one of the 16 beakers contained 4 blue beads, that estimate would accord with the assumption that all the beads in the urn were blue. On the other hand, if a subject stated that half of the beakers did not contain a single blue bead and half contained 4 blue beads each, this would be a most unlikely outcome on any conceivable assumption about the contents of the urn: one subject gave these values. We do not know whether the subjects make any explicit assumptions about the contents of the urn, but we may perhaps infer, from the number of blue and yellow beads indirectly assigned by each subject, that they do tend to assume approximately equal proportions. Nevertheless, few of the individual distributions have a high probability of arising in terms of the number of blue and yellow beads implicitly assigned.

In order to assess the effect on the distributions given of increasing the amount of information about the contents of the urn, we have carried out an experiment with a comparable age group, 10–12 years ($N = 87$), in which the subjects are told that there are equal numbers of blue and yellow beads in the urn. The resulting group distributions are much the same as those obtained here. We have also studied the case in which the subjects are told that the proportions are 3:1 and there the distributions are very different, especially as given by the 14+ group. A detailed analysis of these two experiments will be given in a separate paper.

Secondly, we can inquire whether the subjects at the ages 10+, 12+ and 14+ have any preference for assigning equal values to 0B and 4B and/or equal values to 1B and 3B, by comparing the actual values assigned to these categories with the values assigned to the remaining categories, taking them in pairs. This comparison is shown in Table 6.

Table 6. *Categories of beakers containing* 0, 1, 2, 3 *or* 4 *blue beads compared in pairs of category*

(N = 87 (10+ years) + 31 (12+ years) + 26 (14+ years).)

Paired categories		Proportions of samples assigning equal values to pairs of categories of beaker			
		10+	12+	14+	Total
0	1	0·17	0·16	0·12	0·16
0	2	0·17	0·13	0·08	0·15
0	3	0·22	0·10	0·19	0·19
0	4	0·16	0·42	0·42	0·26
1	2	0·15	0·23	0·23	0·18
1	3	0·17	0·45	0·35	0·26
1	4	0·26	0·10	0·12	0·20
2	3	0·15	0·16	0·23	0·17
2	4	0·25	0·13	0·12	0·20
3	4	0·15	0·13	0·12	0·14

We can calculate χ^2 from the observed number of instances in which equal and unequal values are assigned by the 12+ and 14+ groups to 0, 4 and/or 1, 3 as compared with the number of equal and unequal values assigned to the remaining pairs. $\chi^2 = 37{\cdot}4$, D.F. = 1, $P < 0{\cdot}001$. This tendency to assign equal values to 0, 4 and to 1, 3 does not appear in the 10+ group (see Table 6).

The greater tendency of the older children (12+ and 14+) as compared with the younger ones (10+), to assign equal values to 0, 4 and 1, 3 is indicated by the significance of the χ^2 calculated from the values assigned by the respective groups. This calculation is made by comparing the number of times the different age groups allot equal values to the 0, 4 and 1, 3 categories as compared with all other possible pairs of categories. $\chi^2 = 18{\cdot}7$, D.F. = 1, $P < 0{\cdot}001$.

This effect is brought out clearly in the protocols of three subjects of a subsample of 30:

(1) Age 12 years 11 months:

The beakers with one yellow, and 3 blue, and those with 3 yellow and one blue, have opposite amounts so I've made them the same. And I've done the same as that with the ones with 4 blue and no yellow; and 4 yellow and no blue.

	No. of blue beads				
	0	1	2	3	4
Distribution given	4	2	4	2	4

(2) Age 12 years 7 months:

There will be most in number V (2 B, 2 Y) as there is an even chance for each two. Numbers I (4 B, 0 Y) and II (0 B, 4 Y) rule each other out and so I have put them even. Numbers III (3 B, 1 Y) and IV (1 B, 3 Y) will rule each other out also. As there is not an even number left I simply put 4 for number III (3 B, 1 Y) and 3 for number IV (1 B, 3 Y).[1]

	No. of blue beads				
	0	1	2	3	4
Distribution given	2	3	5	4	2

(3) Age 12 years 10 months:

I thought that the beakers having an equal number of beads in them would not be many and I only put one. Also the first and second answers (0 B, 4 Y and 4 B, 0 Y) are likely to be about the same, and the third and fourth answers (1 B, 3 Y and 3 B, 1 Y).

	No. of blue beads				
	0	1	2	3	4
Distribution given	4	3	1	2	6

Thirdly, bearing in mind that the subjects implicitly allot approximately equal proportions to blue and yellow, it is of interest to see whether they tend to assign higher values to 2 B, 2 Y as compared with any other combination. If so, it might indicate that the samples in the beakers are regarded as miniatures of the contents of the urn and hence be allotted equal numbers of blue and yellow beads. This is supported by the fact that one subject assigned all 16 beakers to the value 2 B, 2 Y. Furthermore, there is a tendency, when the proportions in the urn are specified as 3:1, to allot high values to 3 B, 1 Y.

We can calculate the values assigned by all three age groups to 2 B, 2 Y as compared with the values assigned to 0 B, 4 Y; 1 B, 3 Y; 3 B, 1 Y and 4 B, 0 Y (see Table 7). There is no difference between the age groups in this respect, but 2 B, 2 Y receives the highest value in a significantly large number of instances. $\chi^2 = 9{\cdot}3$, D.F. $= 1$, $P < 0{\cdot}01$.

Table 7. *Number of times highest value is assigned to the different categories of beaker*

Age	No. of blue beads 0	1	2	3	4
10+	8	10	18	9	7
12+	3	1	7	5	5
14+	2	1	5	2	4
Total	13	12	30	16	16

Fourthly, we can see whether there is any significant tendency to assign higher values to 1 as compared with 0 and to 3 as compared with 4. If such higher values were found to be present they might indicate a tendency towards a distribution.

χ^2 calculated from the combined values assigned by all three groups to $1 \gtreqless 0$ and to $3 \lesseqgtr 4$ is $9{\cdot}6$, D.F. $= 1$, $P < 0{\cdot}01$. Inspection of Table 8 shows that in all three groups there is a tendency to assign greater values to 0 than to 1, and to 4 than to 3. Instead of a tendency towards a distribution, we find a tendency away from it. This might be accounted for by the marked preference for balance and the avoidance of asymmetrical arrangements as demonstrated above in relation to Type I.

---

[1] This subject herself numbered the questions I to V.

Table 8. *Comparison between values assigned to categories of beaker containing* 0 *blue and* 1 *blue, and* 3 *blue and* 4 *blue*

	1, 0			3, 4		
Age	>	=	<	>	=	<
10+	32	15	42	27	16	44
12+	8	6	17	10	6	15
14+	10	3	13	9	3	14
Total	50	24	72	46	25	73

VI. DISCUSSION OF RESULTS

Perhaps we should first note the effects on subjects' implied assumptions about the contents of the urn of not specifying the proportions. In all three types of experiments the behaviour of the subjects does not, on the whole, display a great deal of scatter around mean values of 0·5 blue and 0·5 yellow. Although they have an exceedingly wide choice of possible values, some 10% of the subjects chose equal values of blue and yellow beads when asked to state what they thought to be the contents of the urn. Their implicit estimates of the contents of the urn, in all three experiments, as derived from their guesses or distributions, also show this limited range of choice. These implicit estimates when the proportions in the urn are not specified resemble the corresponding estimates made in other experiments by comparable groups of subjects when the proportions in the urn are specified as 1:1. When, however, the proportions are specified as 3:1 the implicit estimates are quite different. We may thus conclude that when the proportions are unspecified there is a tendency for each child to assume that the proportions are equal or nearly so. The assumption of equality, under the conditions described, seems to be the simplest and the most naturally acceptable to the child.

In the three types of experiment the subjects' choices, at least the majority of them, are not such as could have occurred by random selection. In Type I there is the preference for certain kinds of arrangement. In Type II the subjects do not tend to allot similar proportions of blue and yellow beads to the different beakers. In Type III the choice of numbers which add to 16 is not random and there is a preference for 2, 2 and for 0, 4.

So far we have been discussing the group means. Some light is thrown upon the individual choices by the explanations given by a group of 12+ ($N = 30$), although we do not suppose that these explanations fully account for the choices actually given. Some 20% of the group appear to have some idea of a distribution. They mention in their explanations one or more of the reasons for the various effects which have emerged in the analysis. Some 40% produce some definite reason for the manner in which they have made their choices, such as, for example: 'I wanted to have every number different' or 'I did not think any of the numbers would be alike'. The remaining 40% gave no definite reason for their choices, e.g. 'They were the first numbers which came into my head and added up to 16. When I had done the first three I divided what was left into two as I thought best.'

Even at the age 10+ there is a tendency to assign the highest value to 2B 2Y. But the preference for assigning equal values to 0, 4 and to 1, 3 appears only in the age group

12+ and upwards. It does not appear in the 10+ group. This preference may be an important new step beyond the level reached by the 10-year-olds for it enables them to construct a symmetrical distribution. They seem to be applying a 'principle of indifference' in that they assign equal likelihood to 1 B, 3 Y as to 3 B, 1 Y.

There thus appears to emerge, within the age range 10–15 years, four stages in the development of the idea of a distribution. First, there is the belief that the numbers in the distribution will tend to vary. Secondly, there is the general tendency to assume that the category of equal proportions (2 B, 2 Y) will occur most often, although a few children believe that it will occur least often. Thirdly, there is the tendency to assign the same likelihood to categories with similar structure (1 B, 3 Y = 3 B, 1 Y; 0 B, 4 Y = 4 B, 0 Y). A further stage, which is apparently not yet reached at age 14+ when the proportions are unspecified, is to assign greater likelihood of occurrence to 1 B, 3 Y than to 0 B, 4 Y and to 3 B, 1 Y than to 4 B, 0 Y.

We may venture the hypothesis that the stages we have distinguished in the development of the idea of a distribution partly represent logical 'structures' (in Piaget's sense) that emerge at different ages and partly represent the fruits of experience. These 'structures' undergo transformations during development. At first they are shaped by purely subjective preferences of one kind or another. These preferences gradually weaken and become displaced by factors inherent in the objective situation.

We should like to thank the Education Authorities at Manchester and Bolton and the principals and staff at their schools for their kind help in making it possible for us to carry out these studies.

(*Manuscript received* 1 *July* 1954)

THE SCIENTIFIC STATUS OF MENTAL MEASUREMENT

By F. W. WARBURTON

Department of Education, University of Manchester

I. *Psychology as an exact science* (pp. 122–124). II. *Scales of measurement in mental testing* (pp. 125–129). *References* (p. 129).

I. Psychology as an exact science

The nature of science is a matter of definition. Despite Kant's rejection of scientific psychology on the ground that mental states cannot be measured, Comte's objection that experimental methods could not be applied in psychology, and the pessimism of William James, psychologists have been in general agreement for over a hundred years that their subject is a science. They also agree that it is not an exact science, although if the exact sciences are defined as those that permit of precise quantitative measurement, it is clear that in certain areas psychology has already laid claim to be exact. The subject is commonly regarded as inexact in the more fundamental sense that there is imperfect agreement between the results of different investigations into the same problem, and that wide variations exist in individual and in group performance under similar experimental conditions.

This paper is largely concerned with the fields of mental measurement and personality testing. Its first purpose is to consider in turn certain factors making for variability of performance in different investigations, and to discuss very briefly the advantages and disadvantages of attempting to bring these factors under control. Six factors are considered: (i) the use of subjective judgements, (ii) control of the experimental situation, (iii) differences between the tests or indicators assumed to represent certain psychological variables, (iv) sampling error, (v) multidimensionality of the data, and (vi) variation within the individual.

The argument leads to a defence of psychometric techniques.

(i) An indispensable condition of an objective experiment is that the readings made by different observers agree. In other fields scientists may dispute the value of certain experiments or the techniques employed, but they do not disagree (in ordinary circumstances) about the actual readings. In statistical terms, inter-judge correlations should be plus one.

An objective psychology defined in this way will exclude two main types of data, first, aesthetic, moral and literary judgements, and secondly, assessments about which the observers fail to agree because the data are too complex to be fully understood. For example Binet testers will sometimes differ in their opinion as to whether a child understands the meaning of a word or the nature of a geometrical shape, or, in a reading scale, the testers may disagree about whether a child has read certain words correctly. The difficulty in these cases is not that the judges carry their own preferences with them, but that it is impossible to obtain an unambiguous answer to the question. The interpretation of many items in projection tests also falls into this category. Over a wide area of psychology, subjectivity of judgement springs from both these causes, the observers applying different standards to equivocal material.

The amount of unreliability in subjective judgements is extremely high. In the cognitive field we find that correlations between examiners usually fall into the 0·5–0·85 range. In personality assessment, inter-judge reliabilities are at much the same level, and average about 0·70. This means that, in terms of variance, a very considerable proportion of the information, often more than half, is due to the experimenter and not to the experiment. Observers also vary considerably in their own judgement from occasion to occasion, which adds to the overall unreliability. Exactitude in the experimental findings cannot be expected if the ability and personality of the investigator enters into the results.

(ii) Another major factor leading to disagreement between the results of different researches in the applied field is the impossibility of controlling the environment in real life situations. Assessments of a boy's personality in school A cannot be directly compared with those made in school B, since conditions in the two schools are different, and each estimate is relative to a different population. This introduces an additional element of unreliability which would be present even if the measures of the boy's personality at school were completely objective, such as the number of attendances at meetings of school societies. This limitation extends over much of the vocational field; it applies to all biographical material, and to follow-up data in the school, university or occupation. In the study of the behaviour of individuals in small groups also, there is no rigid control of the environment, since the members of the group differ from occasion to occasion. The same person may be loquacious in one group and silent in another. Greater agreement between the results of different investigations should come with increased standardization of the environment. It is well known, for example, that follow-up studies in the vocational field show greater agreement if the criterion is success on a training course, than if follow-up is made into success in a job. There should be still greater agreement if the training course is the same for all candidates, and so on. The only branches of psychology which seem likely to achieve exactness comparable with that found in the older sciences are accordingly those confined to the laboratory and a controlled environment, i.e. physiological, experimental and animal psychology, mental measurement, objective tests of personality and laboratory studies in social psychology. Together they constitute a considerable body of knowledge. As long ago as the eighteen-nineties Dilthey (see, for example, Hodges, 1952) suggested the division of psychology into two parts, an explanatory psychology that would imitate the natural sciences and attempt to discover the elements of conscious experience and the laws of their combination, and a descriptive psychology, which is an artistic process in which we are constantly aware of the relation of the parts to the totality. Many psychologists would regard any such distinction as artificial and an impoverishment of the subject. They take the view that psychology should not be divided according to the methods of analysis employed, but according to content, its data falling naturally into a few main branches, despite overlapping at the borderline.

(iii) A third source of inexactitude in psychological research is the use of different, and sometimes very small, samples of subjects to investigate the same problems. Experiments designed to reveal the influence of major factors such as sex and age may, in fact, yield conclusions which are true only in certain geographical areas or in certain social strata. This is another problem which is insoluble in the educational and vocational fields, although a greater attention to sampling differences both in the design of experiments and in the evaluation of the significance of the results, would help to bring more uniformity into research findings.

(iv) A fourth source of inaccuracy lies in the name or 'indication' given to variables. The Matrices and the Dominoes tests, for example, are both assumed to indicate non-verbal intelligence, although the correlation between the two tests is considerably less than their reliabilities. In the personality field, different tests of the same trait often show very poor agreement, correlating at the 0·4 level. In some researches there is less relation between the same traits than between nominally different traits. From this point of view, factor analysis can be used, not to identify groups of similar tests, but to indicate the best single 'marker' i.e. the most typical test in a given field. These marker tests could be used in important investigations, non-verbal intelligence being replaced, let us say, by Matrices ability. Something of this sort happens by common usage when it is agreed that a certain measure will best serve our purpose, such as the Binet test in child guidance, the Wechsler-Bellevue in adult work, the Matrices in non-verbal intelligence, and the T.A.T. and Rorschach projection tests. Increased agreement between empirical results is likely to follow from agreement between the tests employed. These further conditions imply an even wider separation between a more exact psychology and everyday behaviour. The applied psychologist is building a mechanical man, whose characteristics are Matrix plus Stenquist plus Strong plus *X*.

(v) Another difficulty arises from the complexity of the subject-matter, the multidimensionality of the data studied by the psychologist is likely to result in further disagreement between the empirical findings. Guttman's scale analysis provides a valuable technique for isolating single dimensions, but once again the difficulty lies in extracting such dimensions from complex variables.

(vi) Variation in individual and group performance. This paper is concerned with differences in the results of different investigations, and not with test-retest reliability. It is well known, however, that variation in performance occurs even during a testing session. This factor will add to the disparity between the experimental results, just as snakes are more difficult to measure than carpets.

It is not suggested that exactness in the sense of precise agreement between the results of different experiments will come with the type of variable at present used in mental testing, even when identical, objective, unidimensional tests are given to representative samples in controlled situations. It is true that variation due to differences between judges and between tests may be removed by the use of identical objective tests. On the other hand, the control of the environment, of the experimental sample, and of the number of dimensions cannot be complete. Variation which remains after these rigid experimental controls have been applied falls into a different category. We do not seek to reduce this variation, but to gain an exact knowledge of its limits. It is for this reason that the theory of probability plays a large part in statistical psychology. If everyday behaviour is ever resolved in terms of static unfluctuating measures, it seems unlikely that they will be closely related to real life variables such as intelligence and dominance, although they might be physiological terms or resemble some of the constructs used in learning theory. The price to be paid for increased stability in experimental results is withdrawal from life, and opinion will differ as to whether the game is worth the candle. In any case, it is not to be expected that the postulates of measurement appropriate to the older sciences are necessarily relevant to psychology, in which the subject-matter is more complex. The primary aim of the psychometrist is not to reproduce the type of scale used in the physical sciences. Human beings do not behave like physical objects and it would be surprising if the same methods of measurement were efficient in both cases.

II. SCALES OF MEASUREMENT IN MENTAL TESTING

The remainder of this paper consists in a reply to certain conventional criticism of the measures used in applied psychology, the critics taking the view that scores derived from mental tests are invalid since some of the fundamental principles of measurement are broken by psychometric techniques. These principles include the law of transitivity, which states that if *A* is greater than *B* and *B* is greater than *C*, then *A* is greater than *C*. In terms of the difficulty of test items this is not true for all individuals. A person can usually answer correctly some questions which are more difficult than those in which he fails. It is as if the same person were less than 5 ft. 8 in. tall but more than 5 ft. 9 in. The root difficulty arises from the complexity of the subject-matter. We are forced to measure multidimensional data in terms of unidimensional scales. Let us first of all take an example of a single dimension, which is measured in terms of a single scale (Table 1). Four persons are given a test comprising four questions: (i) Are you more than 5 ft. 4 in. tall? (ii) Are you more than 5 ft. 5 in. tall? (iii) Are you more than 5 ft. 6 in. tall? (iv) Are you more than 5 ft. 7 in. tall? These questions are plotted in order of difficulty and the persons in order of merit as follows:

Table 1

	Question				Total
Person	64 in.	65 in.	66 in.	67 in.	score
A	✓	✓	✓	✓	4
B	✓	✓	✓	×	3
C	✓	✓	×	×	2
D	✓	×	×	×	1
Total persons (difficulty)	4	3	2	1	

A answers 'Yes' to the most difficult question, i.e. he is more than 67 in. tall. Consequently, he also answers 'Yes' to the easier questions, since he is also more than 66, 65 and 64 in. Similarly, although *B* is not more than 67 in. tall, he is more than 66 in. and therefore also more than 65 and 64 in. and so on. The ticks and crosses in the table form solid blocks.

On the other hand, in Table 2, in which the answer plot resembles those obtained for mental tests, person *B* can answer question *S* correctly, but fails in question *R* which is less difficult. The intransitivity of the answer pattern follows necessarily from the multidimensionality of the data. To illustrate this, let us suppose that we are trying to measure a multidimensional variable, say a person's shape, which depends on height and breadth, in terms of a single dimension, the ability to pass through a series of doors. The test comprises a set of four doors, some narrow, some squat, some squarish, and so on, labelled *P*, *Q*, *R*, *S*, respectively. Let us also suppose that, like the mental tester, we do not know the exact nature of what we are measuring, although we are able to devise questions that appear relevant. We cannot distinguish between the height or breadth of either persons or doors, we simply know that different persons have different shapes and can pass through certain doors. If the questions are arranged in order of difficulty and the persons in order of merit we have the type of answer pattern shown in Table 2. The most difficult question is door *S*, since only one person can pass it. The easiest question is door *P*, which is passed by all four subjects.

Person *B* is tall and thin, so thin, in fact, that he is the only person who can squeeze through the narrow and difficult door *S*. On the other hand, his height prevents him from getting through door *R*, which is easier than door *S*. Person D is so fat that he is the only one who is unable to squeeze through the easy door *Q*, although he can get through the more difficult door *R* because he is short, and so on. The difficulty is that we are trying to reduce a multidimensional to a unidimensional scale, to express two variables in terms of one to measure height and width in terms of doors only. This type of answer plot in which the ticks and crosses are arranged higgledy-piggledy, is to be expected in mental tests.

Table 2

	Question (door)				
Person	*P*	*Q*	*R*	*S*	Total score
A	✓	✓	✓	×	3
B	✓	✓	×	✓	3
C	✓	✓	×	×	2
D	✓	×	✓	×	2
Total persons (difficulty)	4	3	2	1	

Another explanation may be advanced of this type of answer plot. It may be that the shapes of the subjects change while the experiment is in progress. They might, for example not be human beings at all, but amoeba. Consequently, a question or door which is difficult to a certain individual one moment may be easy the next. There is no way of telling from the answer pattern alone whether this unevenness is mainly due to the multidimensionality of the data or to the short-term variability of the individual.

The point of view taken in this paper is that the intransitivity of test measures, which comes largely from the use of multidimensional data, is unavoidable, regrettable, and of little consequence. In fact, intransitivity might well be handled satisfactorily by a branch of mathematics, such as the theory of groups, in which the formal requirements are less rigorous.

Two further criticisms of mental testing based on the theory of measurement are (i) that if measurement is used in psychology it should be in terms of linear units, which are constant in size, whereas mental tests are scored in terms of items of different difficulty, and (ii) that there is no true zero in mental measurement so that it is impossible to use ratios: the statement, for example, that one person is twice as intelligent as another has no meaning.

Various types of scale are possible, of which the most important to psychologists are: (i) ordinal scales, such as the scale of hardness for minerals; (ii) interval scales, which demand equality of intervals but do not require an absolute zero (Fahrenheit temperature scales are examples of this type of scale, in which ratios cannot be compared, e.g. a temperature of 212° F. is not twice as hot as a temperature of 106° F.); (iii) ratio scales, which demand equality of intervals and an absolute zero. Ratios may be compared, e.g. a person weighing 16 stone is twice as heavy as a person weighing 8 stone.

In psychology, interval and ratio scales play an important part in experimental and animal psychology. Reaction times, sensory thresholds, maze learning times and measures of fatigue, are chiefly of interest as absolute rather than ordinal measures. In vocational psychology, too, some occupational assessments such as size of vocabulary and shorthand speed are usually regarded as absolute rather than relative. For example, the number of words per minute that a shorthand writer can take down is more important than her position in the commercial school.

The important point is that in the fields of mental measurement, personality testing and experimental social psychology, scales are rare in which some form of rank order is not clearly implied. Some measures of ability, such as I.Q.'s may become sufficiently well known to appear to be absolute, but an I.Q. of 160, like a size 17 shoe or a golfing score of 70, is relative to a population. We know from experience that it is uncommon. We rank it.

The applied psychologist is concerned with the distance between people in society, not between objects in space, he is more interested in rank order than in absolutes, and he therefore changes his raw scores into quotients, percentile ranks, mental ages, and scales based on the standard deviation, which indicate the relation of each measurement to other measurements in the same universe. The zero point is in the middle of the scale; it is the average score, and the scale extends on either side of it. To have 'zero' intelligence, in the ordinary sense of the term 'zero' is to be dead, and beyond the reach of the mental test. It is possible to build a scale of ability in terms of difficulty, which, although not unidimensional, has many of the same characteristics as the ordinary measuring rod, with a zero point and units which are additive and equal (Peel, 1949). Such scales are rarely constructed because they are not particularly useful. It would not be of outstanding importance (unless the measures were stable) if a cerebrometer were invented which measured absolute intelligence in equal and additive units of cerebral plasticity. The next step would still be to change the readings into some form of ordinal units. Critics of psychological measurement who insist on the use of interval and ratio scales might as reasonably criticize a geologist because he does not know the percentile score of a pebble that weighs one and a half ounces. The geologist would reply that he is not interested in the relationship between individual pebbles in a society of stones. This is particularly true in the field of personality, in which no universally accepted linear scale has been established, even by common usage. In the course of an account of the limitations of mental testing Chambers (1943) remarks that persons have been asked to give marks to their emotional moods on a scale ranging from zero to ten, and so to arrange the marks that they form a normal distribution. He asks whether this is so very far removed from the absurdity of treating love as a commodity which may be weighed and made into parcels and tied up with pink ribbon. It is an interesting commentary on the appropriate scale to use in this situation that when a young woman asks her young man how much he loves her, and he answers three and a half, or two million and two, his reply is accepted as persiflage, but when the discussion turns to rank order, things become grimmer. On hearing that he is not placed at the hundredth percentile, the young man may regard this scale as real enough, despite the fact that it is not based on linear and additive units, to take his life. The primary consideration is, in fact, not the type of scale used. The high jump and the mile race are measured in the ratio scales of space and time used by the physicist, but the underlying aptitudes are complex and the problem of human variability remains.

In short, in our present stage of knowledge, the variables used by the applied psychologist are multidimensional. It is difficult to think of any human characteristic, from religious conversion to the knee jerk, which is not multidimensional, and this limitation has to be accepted. We also have to accept the fact that human behaviour, at the level studied by the psychologist, fluctuates from occasion to occasion. It often happens, for example, that a person is, according to examination results, a good mathematician one day and a poor one the next; an introvert at breakfast and an extrovert at dinner; his performance in the high jump may fluctuate within a single minute. This variability extends over the whole range of behaviour, and is unavoidable. Applied psychologists have to define their variables in terms of ratings and test scores, and to use these scores as their starting-point without making any assumptions about their underlying nature or the number of dimensions of the traits concerned.

Another line of criticism of the methods of measurement used in mental testing is derived from the holistic point of view, that to take over the mathematical methods of the physical sciences and to use them for the purposes of psychological description and analysis commits us to a fundamental flaw. We cannot add units of ability as we add twelve inches to make a foot; nor can we put samples of behaviour into a balance and subtract and multiply them as we subtract and multiply ounces and pounds. The psychologist, according to this view, is not justified in assessing traits and persons in the same way as the physicist measures lengths and weights: the variables with which he is concerned do not differ simply in numerical magnitude, they are qualities, not quantities. He assumes that mental tests are based on factors which can be combined or resolved like forces in accordance with the parallelogram law of dynamics, i.e. that they can be presented by vector addition. Mental characteristics, so the argument runs, cannot be treated in this way: they produce new and unpredictable properties in combination, resembling chemical properties not mechanical forces. An individual does not, it is argued, respond to a certain situation merely by using a particular quality in isolation, he reacts as a unified personality.

A similar criticism is that test scores are a kind of blurred average. Different people may give identical answers to a test, but use different methods of arriving at them. One person may visualize problems diagrammatically, another may rely on considerations of abstract logical relationships, and another fall back on the recollection of similar problems solved in the past. These individuals can be regarded as equal only in so far as they have gained equal marks on a particular test. They cannot be said to possess equal quantities of some such capacity as 'general intelligence', or 'manual dexterity'. The treatment of subjects with equal marks as of equal intelligence cannot, according to this view, be seriously maintained. With regard to this argument, Thurstone (1947, p. 54) considers that if critics were consistent who point out that identical test scores do not necessarily measure the same mental processes, they should also object to the statement that two men have identical incomes, for one of them earns while the other one steals. Burt (1940, p. 137) has pointed out that when the householder receives a bill from his grocer he does not argue that the pleasures of cheese and chocolate are commensurable neither with each other, nor yet with a magnum of champagne, and that consequently the addition exhibited in his bill contains a fundamental flaw.

The limitations of the method of analysis are common to science. In order to understand any situation we analyse it into elements and study the separate elements in isolation. This procedure is essential in any investigation of complex material. When the isolated elements are recombined, something is missing in the reconstructed situation, which was present in the original situation. Physical bodies are incompletely analysed into electrons and protons, and water is described, rather absurdly, as the combination of two gases. This difficulty is particularly marked if the subject-matter is as complex as human behaviour.

The analytic method has proved extremely useful to the applied psychologist, who, whilst stressing the limitations of analysis and the importance of studying the whole personality, also uses analytic concepts such as I.Q., aggression, reading age, distractibility and social status. It is difficult to see how he could proceed in any other fashion. The methods used at present are justified by the results: they have pragmatic sanction. Work in psychometrics, in vocational and educational selection, in social psychology and in the field of personality cannot stop because the methods of measurement are not the same as those used in the older sciences.

REFERENCES

BURT, C. (1940). *The Factors of the Mind*, University of London Press.

CHAMBERS, E. G. (1943). Statistics in psychology and the limitations of the test method. *Brit. J. Psychol.* XXXIII, 189–99.

HODGES, H. A. (1952). *The Philosophy of Wilhelm Dilthey.* London: Routledge and Kegan Paul.

PEEL, E. A. (1949). Item difficulty as the measuring device in objective mental tests. *Brit. J. Stat. Psychol.* II, 69–75.

THURSTONE, L. L. (1947). Multiple Factor Analysis. University of Chicago Press.

BODY-CATHEXIS AND PERSONALITY

By SIDNEY M. JOURARD and PAUL F. SECORD

Department of Psychology, Emory University, Georgia

I. *Introduction* (pp. 130–131). II. *Measurement of body-cathexis* (p. 131). III. *Correlates of body-cathexis* (pp. 131–137): (i) *Body-cathexis, anxiety and security concerning the self* (pp. 131–132); (ii) *An interpersonal correlate of body-cathexis* (pp. 132–133); (iii) *Body size and body-cathexis* (pp. 133–134); (iv) *Body-cathexis and the ideal female figure* (pp. 134–135); (v) *Demonstration that the concept of ideal figure is shared* (pp. 135–136); (vi) *Sex differences in differentiation of body-image and self-concept* (pp. 136–137). IV. *Discussion* (p. 137). V. *Summary and conclusion* (p. 137). *References* (p. 138).

I. Introduction

The body has been studied by psychologists from a number of different viewpoints, but a curious omission has been the study of *feelings* about the body.

There can be little doubt that feelings about the body have marked behavioural consequences, as both casual and clinical observation attest. Thus, in our society one can readily note the amount of care and attention that persons devote to the grooming and modification of their body structure and appearance. From a clinical viewpoint it is a common observation that emotionally disturbed persons express concern about the appearance and functions of their body and its parts. Yet, reports of systematic empirical research into feelings about the body are rare.

An early interest in feelings about the body was expressed by the psychotherapists. Freud and Adler called attention to feelings of narcissism, anxiety, and inferiority, with the body and its parts as object. Schilder (1935) devoted a monograph to the role of the 'body image' in personality functioning and adjustment. But the method of inquiry followed by these workers was primitive; they made extensive generalizations about personality-in-general from the self-observations of patients in therapy.

Perhaps the earliest attempt to bring more sophisticated psychological method to bear on body feeling is provided by Schilder (1938). In his book *Psychotherapy*, he describes some questionnaires developed by him as an adjunct to group and individual therapy. These questionnaires served as a more efficient means of exploring the beliefs, feelings, and memories of patients undertaking therapy than the free interview. Schilder was interested in these data for reasons stemming from his theoretical concern with the body-image.

Machover (1948) has provided a 'projective' method of studying the body-image which utilizes drawings of the human figure. This method may be useful as a source of hypotheses concerning the single case, or as a method of confirming hypotheses deriving from other approaches, but it has not proven of much value for more systematic study.

Secord's word association procedure (1953) may be viewed as an improvement in methodology. He devised a set of 75 homonyms which could be perceived as words relating to the body, or as non-body-related words. The subjects were required to respond to the oral presentation of these words by writing down the first association that occurred to them. The words *colon*, *graft*, and *tablet* are illustrations. Three bodily responses might be: *colon*:*intestine*, *graft*:*skin*, *tablet*:*aspirin*; while three non-bodily responses might be: *colon*:*comma*, *graft*:*politics*, or *tablet*:*paper*. The total number of bodily responses for the

75 homonyms constituted the subject's score. Secord interpreted the number of body-related responses as indices of body-preoccupation, either narcissistic or anxious, and offered some tentative supporting evidence.

The present writers approach the problem of feelings about the body more directly. They raised the question: 'Can feelings about the body be quantified?' The problem seemed to lend itself nicely to methods already familiar in questionnaires and rating scales. Accordingly, they defined one variety of feeling, body-cathexis—the degree of satisfaction with the parts of the body—and formulated a questionnaire for its direct measurement. Parenthetically, it should be noted that the present use of the term differs from the psychoanalytic meaning which indicates the amount of instinctual energy invested in an object. In the present paper, the term refers to a continuum of feeling ranging from strong positive to strong negative.

II. THE MEASUREMENT OF BODY-CATHEXIS

The most recent Body-cathexis (BC) scale is a listing of forty parts or aspects of the body which are rated by subjects according to the following five-point scale:

(1) Have strong positive feelings.
(2) Have moderate positive feelings.
(3) Have no feeling one way or the other.
(4) Have moderate negative feelings.
(5) Have strong negative feelings.

Thus, if a person were very dissatisfied with his weight, he would encircle a (5).

The forty body parts are: hair, facial complexion, appetite, hands, distribution of hair (over body), nose, physical stamina, elimination, muscular strength, waist, energy level, back, ears, age, chin, body build, profile, height, keenness of senses, tolerance for pain, width of shoulders, arms, chest (or breasts), appearance of eyes, digestion, hips, resistance to illness, legs, appearance of teeth, sex drive, feet, sleep, voice, health, sex activities, knees, posture, face, weight, sex organs.

Although it is subject to some of the pitfalls and shortcomings already discovered in questionnaires or self-inventories this method of study has the advantage of relative simplicity, reliability and much previous work done with it.

In some of the studies to be reported, a self-cathexis (SC) scale similar in nature to the BC scale was employed. Self items ranged from 40 to 56 in the various forms of this scale. Examples of self-cathexis items are: morals, ability to express self, artistic talents, intelligence level, conscience, personality, self-discipline, vocabulary.

III. CORRELATES OF BODY-CATHEXIS

(i) *Body-cathexis, anxiety and security concerning the self*

The first study (Secord & Jourard, 1953) undertaken with this instrument (hereafter called the BC scale for body-cathexis) was conducted with a total of 126 college male and female subjects. The BC scale was submitted to 70 males and 56 females along with a 56-item SC scale, and the homonym test of anxious body concern (Secord, 1953). Some cases were eliminated for reasons of 'response-set'. All three scales were demonstrated to possess adequate reliability, even with the exclusion of these cases.

The following results were obtained, as shown in Table 1.

Table 1. *Intercorrelations among BC-SC scores, the homonym test, and the Maslow test*

Test	Body-cathexis	Self-cathexis	Anxiety-indicator
Honomym test			
Males ($N=45$)	−0·18	−0·15	−0·37*
Females ($N=43$)	−0·41**	−0·23	−0·40**
Maslow test ($N=46$)	−0·37*	−0·52**	−0·41**
Self-cathexis			
Males	0·58**		
Females	0·66**		

* 0·02 level of significance. ** 0·01 level of significance.

Significant correlations were obtained between BC and SC for both sexes, between the homonym test and total BC for females but not for males; and for both sexes between the homonym test and an eleven-item '*anxiety-indicator score*' derived from the BC scale.[1]

A subsequent study with a mixed group of 47 males and females yielded a significant correlation between the Maslow test of Psychological Security-Insecurity (Maslow, Hirsch, Stein & Honigmann, 1945), and BC, SC, and the 'anxiety-indicator score'. These data also appear in Table 1. The Maslow test is of the self-inventory variety, containing 50 items to be answered by *yes*, *no*, or *?*. Sample items are: 'Do you ordinarily like to be with people rather than alone?' 'Do you lack self-confidence?' These items were selected by Maslow from a large number on the basis of the extent to which they discriminated between normals and between persons characterized as insecure on the basis of clinical interviews.

From the results in Table 1 it is clear that body-cathexis is a variable which can be measured, and which is meaningfully related to other variables with an adjustment connotation, namely self-cathexis (the degree of satisfaction-dissatisfaction with aspects of the self), security with respect to the self (as measured by Maslow's scale), and Secord's projective-homonym test of anxious body-concern.

(ii) *An interpersonal correlate of body-cathexis*

The above study demonstrated that different degrees of body-cathexis 'make a difference'; i.e. that BC is indeed a 'significant' variable in personality research and theory. The next research to be described was concerned with the question: 'Will a person's attitudes and feelings about his body be related to attitudes believed to be held by "significant others"?'

Remy (1953), a student of the authors', conducted a study on this problem. Using 48 college males and 51 college females as subjects, he had them fill out three BC-SC questionnaires: first, according to their own feelings; secondly, as they believed their mothers felt about their (the S's) bodies and selves; and thirdly, as they believed their fathers felt. Finally, the subjects filled out a Maslow scale.

The main results of this study are summarized in Table 2. It may be seen that *perceived* mother and father attitudes about the S's bodies and selves correlate to a significant degree with self-evaluations, and that both parental cathexes, and BC-SC scores correlate with the Maslow test of security (Mas).

---

[1] The 'anxiety-indicator' score was based on the sum of the cathexis-ratings assigned by each subject to the eleven items that received the greatest number of negative ratings by the group.

Table 2. *Intercorrelations among BC-SC scores, perceived parental cathexes, and security scores*

	Self-rating		Perceived mother-ratings		Perceived father-ratings	
Test	BC	SC	BC	SC	BC	SC
BC	—	0·68**	0·68**	—	0·56**	—
	—	0·84**	0·74**	—	0·68**	—
SC	—	—	—	0·77**	—	0·66**
	—	—	—	0·70**	—	0·65**
Mas	− 0·50**	− 0·66**	− 0·57**	− 0·63**	− 0·42**	− 0·49**
	− 0·37*	− 0·63*	− 0·32*	− 0·37*	− 0·33*	− 0·37*

The first r in each cell is for $N = 51$ females; the second r is for $N = 48$ males.
* Significant at the 0·05 level of confidence.
** Significant at the 0·01 level of confidence.

The data obtained by Remy signify that within the 'psychological world' of the subjects, if they believe that their parents dislike assorted aspects of body and self, then they hold mainly similar body-self attitudes, and appear to be generally insecure, in so far as the latter is tested by Maslow's test. If one might assume that the subjects are able to estimate their parent's feelings with some accuracy, certain theoretical propositions outlined by Sullivan (1947), Murphy (1947) and Rogers (1951), with respect to the correlation of self-appraisals with the appraisals of significant others are supported by the data. Some knowledge concerning the accuracy of the estimates is needed, however, before any definite conclusions can be drawn.

(iii) *Body size and body-cathexis*

The next studies to be reported are concerned with the question: 'Why will a subject report that he likes or dislikes a given body part?' Two separate researches support the contention that the *measured size* of certain body aspects is an important determiner of the subject's feelings about those body parts.

The first of these studies (Jourard & Secord, 1954) was conducted with 62 college males. Subjects first filled out the BC-SC scale. Without prior warning, measurements were then taken of height, weight, width of shoulders, circumference of chest when relaxed and expanded, and circumference of biceps when relaxed and contracted. Correlations were computed between these measurements and the five relevant BC ratings: *height*, *weight*, *shoulder width*, *chest* and *muscular strength* (see Table 3).

Table 3. *Correlations between size and cathexis for selected body aspects*
($N = 62$ males.)

Body-cathexis	Actual size	r
Height	Height	− 0·55**
Weight	Weight	− 0·04
Shoulders	Width of shoulders	− 0·36**
Chest	Chest (relaxed)	− 0·33**
Chest	Chest (expanded)	− 0·37**
Muscular strength	Biceps (relaxed)	− 0·52**
Muscular strength	Biceps (contracted)	− 0·54**

** $r = 0·33$ at the 0·01 level for D.F. = 60.

All of the correlations, except that with weight, while low, are significant. They indicate, from their direction, that *large* size of the relevant body parts is associated with positive cathexis, while the reverse is true for small size. It was suggested in the report of this research that these data reflect the cultural stereotype or ideal with respect to masculinity; namely, that in our culture, *it is important and masculine to be big.*

The reverse trend was found with females (Jourard & Secord, 1955). Cathexis-ratings on a seven-point scale and direct measurements were taken of the following twelve body parts of 60 college females: *height, weight, bust, waist, hips, thighs, calves, ankles, feet, nose length, shoulder width* and *neck length.* Correlations between these pairs of measurements are shown in Table 4. All but four of these correlations between size and cathexis ratings were significant. The direction of these *r*'s suggested that *small size* was most desirable to the subjects, with the exception of *bust*, where large size was desired.

Table 4. *Correlations between cathexis-ratings and measured size of body parts*

($N = 60$ females)

Height	0·19	Calves	0·39**
Weight	0·37**	Ankles	0·46**
Bust	−0·25*	Feet	0·52**
Waist	0·32*	Nose length	0·22
Hips	0·43**	Shoulders	−0·13
Thighs	0·42**	Neck length	−0·17

* Significant at the 0·05 level. ** Significant at the 0·01 level.

The findings with males and females both implied, without demonstrating, the presence of cultural ideals with respect to masculinity and femininity. This ideal was not studied among the males, but was explored extensively among the group of sixty females described above.

(iv) *Body-cathexis and the ideal female figure*

The sixty females reported above were asked to give *estimates* of the size of the twelve body parts, and also to specify the dimensions which they believed would be most ideal for them. The hypothesis governing this aspect of the study may be stated as follows: cathexis for a body part is a function of conformity with or deviation from the subjective ideal; the more closely a girl's body approximates her concept of ideal dimensions, the more she will be satisfied with this body part.

Accordingly, deviation scores were computed for the five body parts most likely to be known by the *S*'s; viz.: *height, weight, bust, waist* and *hips.* Two deviation scores were obtained, one, by subtracting the actual measurements from the ideal dimensions, ignoring sign, and another, by subtracting the estimated measurements from the ideal. These may be regarded as measures of objective, or 'real' deviation from the ideal, and perceived deviation from ideal, respectively. As shown in Table 5, both sets of scores correlate quite highly with cathexis-ratings for the relevant body parts.

Table 5. *Correlations between cathexis-ratings and deviations from ideal size for selected body parts*

(The two discrepancy scores are obtained by subtracting the measured size and estimated size, respectively, from self-ratings of ideal size ($N = 60$ females).)

Discrepancy score	Height	Weight	Bust	Waist	Hips
Ideal-measured size	0·41**	0·50**	0·11	0·42**	0·32*
Ideal-estimated size	0·55**	0·61**	0·21	0·50**	0·49**

* Significant at the 0·05 level. ** Significant at the 0·01 level.

Another observation may be made with respect to the nature of the ideal—an observation consistent with the correlations shown in Table 4. In three body aspects (*weight*, *waist* and *hips*) the self-rated ideal is significantly smaller than the measured size; the ideal for *bust*, on the other hand, is significantly larger than the average bust measurement for the group. The data on which these observations are based are shown in Table 6.

Table 6. *Comparisons of mean difference between measured (and estimated) size and ideal size*

($N = 60$ females)

Body part	Measured size	Ideal size	Mean diff.	$\sigma_{\text{mean diff.}}$	C.R.	P
Height	65·69 in.	65·53 in.	0·16	0·20	0·80	>·05
	Est. *65·52 in.*	—	*0·01*	—	—	*>·05*
Weight	125·42 lb.	122·48 lb.	2·94	1·49	1·98	·05
	Est. *126·97 lb.*	—	*4·49*	*1·35*	*3·32*	*<·01*
Bust	34·14 in.	34·83 in.	0·69	0·20	3·45	<·001
	Est. *34·28 in.*	—	*0·55*	*0·19*	*2·89*	<·01
Waist	25·45 in.	24·27 in.	1·18	0·18	6·56	<·001
	Est. *25·57 in.*	—	*1·30*	*0·18*	*7·22*	*<·001*
Hips	37·43 in.	35·06 in.	2·37	0·29	8·17	<·001
	Est. *36·84 in.*	—	*1·78*	*0·23*	*7·74*	*<·001*

Est. = estimated size.

The average deviations from ideal measurements may be described in terms of a hypothetical average girl (in the sample). She is 2·94 lb. overweight, has a bust measurement which is 0·69 in. too small, a waist that is 1·18 in. too big, and hips that are oversize by 2·37 in.

These differences between measured and ideal mean sizes are small. Yet, when we are dealing with body measurements, even these slight differences are highly significant; phenomenologically, the small differences truly make a difference.

(v) *Demonstration that the concept of ideal figure is shared*

Each girl in the study being reported was asked to specify the measurements which would be ideal *for her*. If these ideal measurements were unshared, we might expect them to vary within the sample as widely as direct measurements of body parts; i.e. there would be as many different ideals as there were objective body proportions. If, on the other hand, these ideal concepts are shared as a sort of cultural stereotype, we would expect the variation among ideal measurements to be smaller than that for direct measures. Accordingly, standard deviations were computed for each ideal measure, and these were compared with standard deviations for measurements of the five body parts. As shown in Table 7, the standard deviations for ideal measurements are all significantly smaller than

Table 7. *Comparison of* S.D.'s *for measured size with* S.D.'s *for ideal size*

($N = 60$ females.)

	Standard deviations (S.D.'s)						
Body parts	Measured size (x)	Ideal size (y)	r_{xy}	$\sigma_x - \sigma_y$	$\sigma_{\text{diff.}}$	C.R.	P
Height	2·12	1·63	0·68	0·49	0·18	2·72	<0·01
Weight	16·79	9·17	0·76	7·62	1·25	6·10	<0·001
Bust	2·06	1·31	0·61	0·75	0·18	4·17	<0·001
Waist	1·71	1·15	0·62	0·56	0·15	3·73	<0·001
Hips	2·20	1·56	0·33	0·64	0·23	2·78	<0·01

those for measurements of the actual size of the body parts. Apparently some factor, probably the culturally defined and shared concept of the ideal, restricts the range of variation of acceptable and desirable body size to a much smaller margin than, say, appetite and given bodily endowment allow. This finding surely has implications for personal adjustment, for it signifies that the cultural ideal of female bodily beauty is one that is difficult, if not impossible to attain; thus, many females are doomed to insecurity by the very restrictive nature of the ideal.

(vi) *Sex differences in differentiation of body-image and self-concept*

An interesting question concerns the extent to which an individual assigns similar ratings to all body parts, or, on the other hand, varies widely in rating different body parts. Thus, at the extremes, he may have a 'global' or a 'differentiated' concept of body and self.

Remy (1953) conducted some exploratory analyses of differentiation of the body-image and the self-concept in which he computed individual variances for BC and SC ratings. He found that females give a higher mean variance than males in responding to the BC scale, but the sexes did not differ in mean variance for SC ratings. The males also had a significantly lower mean variance for BC ratings than they did for SC ratings, while the women did not differ in differentiation of body and self.

The relevant analyses and comparisons are shown in Table 8.

Table 8. *Difference between mean variances for self-rated body-cathexis and self-cathexis in male and female subjects*

($N = 48$ males and 51 females.)

	Between-sex comparisons			
	Mean variance			
Scales compared	Males	Females	Difference	*t*
Body-cathexis	0·76	0·96	−0·20	2·13*
Self-cathexis	1·01	1·02	−0·01	—

	Within-sex comparisons		
	Mean variance	Difference	*t*
Males			
Body-cathexis	0·76	−0·25	3·37**
Self-cathexis	1·01		
Females			
Body-cathexis	0·96	−0·06	—
Self-cathexis	1·02		

* Significant at the 0·05 level. ** Significant at the 0·01 level.

These findings suggest that women in our culture make finer cathexis discriminations within their phenomenal body structure than males. Perhaps the greater variability shown by women in assigning cathexis-ratings to their body parts reflects a cultural limitation on the means of attaining security and acceptance from others[1]—a woman's body is somewhat more important to her attainment of acceptance and security than a man's body is to a man. Men in our culture may have more avenues of attaining security through traits of self. The higher correlation among women than among men between

---

[1] Murphy points out (1947, p. 345) that differentiated perceptions emerge from more global ones under the impetus of need-tensions.

BC and security scores which Remy found (see Table 2) is perhaps additional evidence in support of the stated propositions concerning the greater personal and social importance of the female body.

IV. Discussion

From the data reported herein, some general comments about the role of the phenomenal body in personality may be offered.

We can suggest that a person's security—his felt self-acceptability as well as his acceptability to significant others—is related in some way to his feelings about his body. Perhaps the explanation of this relationship is that a person in our culture perceives his body appearance as a tool for impressing others, whose approval he has come to need for his self esteem. If his body does not correspond with his standards (internalized from others), he will feel insecure.

This interpretation is supported by the findings with respect to body size and body-cathexis. Men and women alike, it appears, acquire a concept of ideal body proportions, and vary their cathexis ratings with the degree to which they approximate these ideal specifications.

With the data already at hand, these considerations seem to apply most strikingly to females, although there is evidence to indicate that they apply to males as well.

There remain many unsolved problems relating to body-cathexis. Two important ones which have not been studied concern the origins of the body attitudes, and the change of cathexis-ratings of body parts with time.

The method which has been described should, with suitable modifications, be useful in the solution of some of these problems.

V. Summary and conclusion

Body cathexis is defined as the degree of satisfaction reported by an individual for the parts of his body. A method for measuring this variable has been described, and the results of a number of studies reported. The findings to date may be summarized as follows:

1. Body-cathexis measures are correlated with measures of self-cathexis, anxious body-preoccupation, and security concerning the self.
2. Body- and self-cathexis measures are correlated with measures of perceived parental attitudes to the body and the self; parental attitudes to the body and the self are also related to security concerning the self.
3. The measured size of the body is an important determiner of cathexis for the body. With males, large size appears to be correlated with positive cathexis, whereas with females, small size seems to be desirable.
4. Among females, self-rated ideal size is smaller for most body parts than the measured size; one exception is for *bust*, where the ideal size is larger than the measured size.
5. The concept of ideal size, among females, appears to be a shared, or stereotyped concept of the ideal, and it shows much less variability within the group than measured size.
6. Variability in assigning cathexis-ratings to the parts of the body is greater among females than among males. The sexes do not differ in variability in rating the self. Among males, greater variability is shown in rating the self than the body; the women do not manifest this difference.

The implications of these findings for personality adjustment are briefly discussed.

REFERENCES

JOURARD, S. M. & SECORD, P. F. (1954). Body size and body-cathexis. *J. Cons. Psychol.* XVIII, 184.

JOURARD, S. M. & SECORD, P. F. (1955). Body-cathexis and the ideal female figure. *J. Abnorm. (Soc.) Psychol.* (in the Press).

MACHOVER, KAREN (1948). *Personality Projection in the Drawing of the Human Figure.* Springfield: C. C. Thomas.

MASLOW, A. H., HIRSH, E., STEIN, M. & HONIGMANN, I. (1945). A clinically derived test for measuring psychological security-insecurity. *J. Gen. Psychol.* XXXIII, 21–41.

MURPHY, G. (1947). *Personality; a Biosocial Approach to Origins and Structure.* New York: Harper.

REMY, R. (1953). Relationships among Body- and Self-Cathexes, Perceived Parental Cathexes, and Security. Unpublished master's thesis, Emory University.

ROGERS, C. (1951). *Client-Centred Therapy.* Boston: Houghton-Mifflin.

SCHILDER, P. F. (1935). *The Image and Appearance of the Human body.* London: Kegan Paul.

SCHILDER, P. F. (1938). *Psychotherapy.* New York: Norton.

SECORD, P. F. (1953). Objectification of word association procedures by the use of homonyms: a measure of body-cathexis. *J. Personality*, XXI, 479–95.

SECORD, P. F. & JOURARD, S. M. (1953). The appraisal of body-cathexis: Body-cathexis and the self. *J. Cons. Psychol.* XVII, 343–7.

SULLIVAN, H. S. (1947). *Conceptions of Modern Psychiatry.* Washington: Wm. Alanson White Fdn.

(*Manuscript received* 26 *February* 1954)

THE DVORINE PSEUDO-ISOCHROMATIC PLATES[1]

By AGNES CRAWFORD

Department of Psychology, University of Liverpool

In 1944, Israel Dvorine published two volumes of colour plates; vol. I he called *Colour Perception Testing Charts for Ages 7 and Over*—a diagnostic set of pseudo-isochromatic charts, and vol. II he called *Colour Perception Training Charts* (Dvorine, 1944*a*). These were published by the author himself and described as available from the author, at a cost of $25.00 (Andrews, 1948). In the same year and in 1945 and 1946, Dvorine published a series of papers with titles as follows: 'A new diagnostic method of testing and training colour perception' (Dvorine, 1944*b*), 'Reconditioning the colour blind; a case report' (Dvorine, 1944*c*), 'The use of volume two of the Dvorine colour perception testing and training charts for grading colour perception difficulties' (Dvorine, 1945*a*), and 'Improvement in colour vision in twenty cases' (Dvorine, 1946).

Papers with such titles could not long pass unchallenged, nor could his claims for the tests. Elsie Murray reviewed the tests in 1945 (Murray, 1945*a*), and was unfavourably impressed both by the plates and by the claim that the charts could be and ought to be used for training colour perception. Her paper led to a somewhat acrimonious though largely irrelevant reply by Dvorine (Dvorine, 1945*b*) and to further criticism by Murray (Murray, 1945*b*). A few other cursory reviews appeared, but the only psychologist interested in colour vision testing at that time who seems to have used the new test at all was Louise L. Sloan. In a comparative study of some tests of colour vision (Sloan, 1945*a*), she used eleven plates of the Dvorine test and quotes the frequencies of scores made by a group of 100 subjects with normal colour vision and by a group of forty-seven subjects with colour-defective vision. She did not give much information about the test, but seems to have selected the eleven plates which her results showed to be most discriminating between previously classified groups of subjects. The table she gives shows no overlap between the scores obtained on these eleven plates by the colour-normal and the colour-defective subjects.

Dvorine continued to write about the tests (Dvorine, 1948). It was not, however, one of the tests selected by Chapanis (1948) for his comparative study of five tests of colour vision, so that, presumably, he was not impressed by the test, nor by its author's claims.

It is clear that the first edition of the Dvorine test was not orthodox, and Louise L. Sloan seems to have been the only person to take it seriously and to use parts of it in the orthodox way. Now, in 1953, we have what is called a second edition of the test—really, however, a new version of the test.

The first page of this new version of the test consists of a most handsome acknowledgement by Dvorine to Dr Louise L. Sloan, and an admission that, starting from her paper in 1945, he had taken the eleven plates which she found to be highly diagnostic, and had used these as a 'nucleus for the second edition'. He has added some new plates and tested them on groups of colour-normal and colour-defective subjects, retaining only those plates which as he says, give a 'high percentage of normal subjects passing and colour deficients failing this test'. This, of course, is a criterion suggested by Sloan (1945*b*). Dvorine does not say on what criterion, other than the test itself, the subjects were classified as normal or colour deficient, nor how many subjects he used.

---

[1] Second edition, 1953, published by and available from the author at Eataw Place, Baltimore, Md. No price given.

In the 'Instructions', for giving the test, Dvorine describes it as consisting of two parts: Section One contains, he says, 'One demonstration plate (the number 48 in red on a blue background) and 14 plates made up of eight different colour combinations arranged in pairs of similar colours. Section Two contains one demonstration plate (a blue trail on a red background) and 7 plates featuring trails instead of digits...the second section may be used to test pre-school age children and illiterates, or as a corroborative test' for the first section. He goes on to give details about the light in which the test should be given, the distance from the subjects' eyes, and to say that no more than five seconds should be allowed for each plate. These are points on which he was criticized by Murray (1945*a*).

On a third page, Dvorine gives what he calls 'Analysis' of the results of testing. This is now quite orthodox. Three or more failures on the fourteen digit-plates indicate red-green colour deficiency. Plates 6 and 7 are diagnostic for 'protanoid' and 'deuteranoid' types of colour deficiency. Two errors or one error can be allowed because of what he calls 'figure-ground confusions by otherwise normally seeing individuals'. Though he has, he says, studied such errors carefully, he cannot account for them. The same sort of error, in fact, appears on the Ishihara test, as any item analysis will show, and it should not perhaps be dismissed as easily as Dvorine suggests.

There is, however, a third section of the test, which Dvorine does not mention and which presumably remains as a legacy from his unregenerate days. It appears in the sample Record Form as a test of Nomenclature, and consists of a movable disk inside a double page; on the disk there are circles of colour which can be shown in succession through an aperture. The colours are what he calls on the Record Form, 'light and dark tints of Red, Brown, Purple, Yellow, Blue, Green, Gray and Orange'. The colours on the digit plates seem much purer and clearer than those on the colour-naming disks.

The test, obviously, has a more varied history than tests of colour vision usually have. A copy of the new edition was given to me for review, and the only way to reach a justifiable opinion about it was, clearly, to try it out, along with a well-established test, and a criterion test. Accordingly, the Dvorine test, the Ishihara test (Lewis's printing, 9th edition), and Rayleigh equations on an anomaloscope (Crawford, 1951*a*, *b*) were used, with as many subjects as could be mustered in the time available. Seventy-six subjects were obtained, mainly from the first-year course in psychology here, though this number includes a number of known colour-blind individuals who volunteered or were persuaded to come for test. In all, sixty-five subjects (forty-one women and twenty-four men) had normal colour vision on both the Plate tests, and eleven (ten men and one woman) were colour defective on both of these tests. Some of the subjects in the 'normal' group whose matches on the Rayleigh equation were at the extremes of the distribution could perhaps be regarded as very slightly deviant, but they show no overlap on any of the tests with those whom I have grouped as colour blind. Only the woman subject and one of the colour-blind men did not previously know or suspect that their colour vision was at least different from that of other people.

I propose to deal mainly with the results of the Dvorine test here, since the main purpose of the investigation was to find out how good or how bad this new version of the test was.

For most of the subjects, the first part of the test only was used. About one-third of the subjects did the second part of the test—i.e. tracing out the wavy lines or trails—sufficient to show that the results for that section will be very similar to those obtained from the first part. The colour-naming section was omitted altogether, after a few trials. The colours are unsatisfactory, and the disks seem to provide a test of the subjects' knowledge of present-day fashionable colour names, rather than of simple colour naming.

The scores for the sixty-five colour-normal subjects and the eleven colour-defective subjects on both of the pseudo-isochromatic tests are as follows:

Dvorine test

No. of plates read correctly	Frequency	
14	51	
13	10	
12	4	
11	1	(colour anomalous man)
5	1	(colour anomalous woman)
4	1	
1	2	
0	6	
Total	76	

Ishihara test

No. of plates read correctly	Frequency	
20	18	
19	22	
18	10	
17	8	
16	4	borderline scores
15	3	
14	1	(colour anomalous woman)
8	1	
3	1	
0	8	
Total	76	

It will be seen that the Dvorine test compares favourably with the Ishihara test. For this, the score is based on the twenty figure-plates which can be read by those with normal colour vision. There is one subject—anomalous on the anomaloscope matchings—who has a borderline score on the Dvorine test, compared with seven or eight on the Ishihara test, only one of whom was clearly defective on the anomaloscope matchings. No one who is clearly defective on the Ishihara test has a better score on the Dvorine test, but some who are 'normal' on the Dvorine test (and on the anomaloscope matchings) have questionable scores on the Ishihara test.

An item analysis of the distribution of errors on the Dvorine plates shows a slight tendency for the relatively small number of errors—(30 errors by sixty-five normal colour-vision subjects)—not to be evenly distributed over the fourteen plates, but to accumulate on plates, 2, 3, 4 and 6, though only to the extent of 6 or 4 misreadings, instead of just over 2 that might be expected with this number of misreadings. This, however, is not such an uneven distribution as is found in any item analysis of the Ishihara test (Crawford, 1950; or Pickford, 1950). With some of the less markedly defective subjects who would be found in a larger sample, it might be necessary to make some allowance for these more frequently occurring errors, though the figures available are not sufficiently large to show whether this divergence from an expected even spread of errors over all the fourteen plates is statistically significant or not.

Correlation with the Ishihara test cannot be calculated, nor, with distributions of the kind shown here, could any correlation be calculated with the anomaloscope measurements. However, it is clear from the two frequency tables, that there are no gross misclassifications. The eleven subjects at the bottom of the tables are classified as colour blind by both the tests. It seems, from the present evidence that the Dvorine test discriminates between the colour normal (by the criterion of the Rayleigh equation matchings on the anomaloscope) and the colour defective (by the same criterion). Colour normals do obtain scores of 12 or more out of 14; colour defectives score 11 or less. With a larger sample there is, of course, the possibility that some subjects reading eleven plates normally might also give normal matchings on the anomaloscope, but the proportion of the present sample giving this borderline performance is sufficiently small to justify the suggestion that such cases could be tested, for confirmation, by some other test. On the Ishihara, there are eight subjects who would need further testing to be sure whether their colour vision was normal or not.

Dvorine claims that the test will allow a diagnosis of 'protanoid' or 'deuteranoid' colour defect to be made. This, however, is not so clear-cut as he suggests. Out of the eleven colour defective subjects, only three were clearly classified; these were 'protanoids', and both the Ishihara and the anomaloscope matchings support this classification. The eight others could not be clearly classified, either because they failed to read both of the diagnostic figures on both plates 6 and 7, or because they read both figures on these plates, with a slight error in one only, or with some hesitation. The Ishihara classifies some of these as deuteranopes, but the anomaloscope matchings put them as deuteranomalous rather than as deuteranopes. One unclassified subject appeared as a definite protanope on the anomaloscope. In general, it might be said that the classification appears to be effective for well marked protanopes, but not for deuteranopes, and the test does not distinguish the anomalous trichromats from the dichromats. However, it may be that Dvorine is asking too much of a printed pigment test in expecting that it should be able to make this finer classification.

Apart from the failure to live up to this last claim, the new version of the Dvorine test seems to me to be an effective screening test, to separate the colour normal from the colour defective. The relatively small number of people who would give borderline scores (i.e. eleven plates read correctly) compares very favourably with the eight people who even on this small sample, would have borderline scores on the Ishihara test. While it is perhaps satisfactory from the point of view of test construction that eight out of eleven colour-defective subjects should have scores of 1 or 0 on the test, it is a little unsatisfactory from the subject's point of view. They are shown one demonstration plate with a clear figure on it, and asked to find other figures of the same sort, and then presented with a series of fourteen plates not one of which appears to them to have a figure on it. Apart from the feeling of resentment aroused by the notion that some trick is being played on them, it seems to me that this might, especially with younger subjects, arouse considerable anxiety about their 'eyesight' in general. On the method of scoring the Ishihara in terms of plates which normals can read, there are four plates which colour-blind people can read, which are not counted in the score, and there are eight or ten other plates on which the colour-blind subjects can see alternative figures. If the Dvorine test were being widely used, it might be advisable to have available some of these plates from the Ishihara test, to insert at judiciously chosen moments in the testing session, so that the subject is not too puzzled and worried by his apparent failure to do what the examiner is expecting him to do. This, however, is not a problem of colour vision, but a problem of testing technique.

REFERENCES

ANDREWS, T. C. (1948). Some psychological apparatus. A classified bibliography. *Psychol. Monogr.* LXII, 289.

CHAPANIS, A. (1948). A comparative study of five tests of colour vision. *J. Opt. Soc. Amer.* XXXVIII, 626–49.

CRAWFORD, A. (1950). An investigation of colour blindness. Ph.D. Thesis, Library of the University of London.

CRAWFORD, A. (1951*a*). A description of an anomaloscope. *Brit. J. Physiol. Optics*, VIII, 173–5.

CRAWFORD, A. (1951*b*). Rayleigh equations on an anomaloscope. *Brit. J. Physiol. Optics*, VIII, 225–34.

DVORINE, I. (1944*a*). *Colour Perception Testing Charts for Ages 7 and Over* (vol. I), and *Colour Perception Training Charts* (vol. II). Baltimore: The Waverly Press, $25.00.

DVORINE, I. (1944*b*). A new diagnostic method of testing and training colour perception. *Amer. J. Optom.* XXI, 225–35.

DVORINE, I. (1944*c*). Reconditioning the colour blind; a case report. *Amer. J. Optom.* XXI, 508–10.

DVORINE, I. (1945*a*). The use of volume two of the Dvorine colour perception testing and training charts for grading colour perception difficulties. *Amer. J. Optom.* XXII, 478–9.

DVORINE, I. (1945*b*). Author's comments upon the review of colour perception charts. *Amer. J. Psychol.* XLVIII, 397–9.

DVORINE, I. (1946). Improvement in colour vision in twenty cases. *Amer. J. Optom.* XXIII, 302–21.

DVORINE, I. (1948). Dvorine colour discrimination screening test. *Amer. J. Optom.* and *Arch. Amer. Acad. Optom.* XXV, 130–3.

MURRAY, E. (1945*a*). Review of Dvorine's colour perception charts, vols. I and II. *Amer. J. Psychol.* XLVIII, 276–9.

MURRAY, E. (1945*b*). A reply to Dvorine's comments. *Amer. J. Psychol.* XLVIII, 399–402.

PICKFORD, R. W. (1950). An item analysis of the Ishihara Test. *Brit. J. Psychol.* XLI, 52–62.

SLOAN, L. L. (1945*a*). An improved screening test for red-green colour deficiency composed of available pseudo-isochromatic plates. *J. Optical Soc. Amer.* XXXV, 761–6.

SLOAN, L. L. (1945*b*). Two criteria for the selection of colour vision test plates. *J. Opt. Soc. Amer.* XXXV, 723–30.

(*Manuscript received* 2 *August* 1954)

PUBLICATIONS RECENTLY RECEIVED

Method and Theory in Experimental Psychology. By C. E. Osgood. New York and London: Geoffrey Cumberlege, Oxford University Press. 1953. Pp. viii + 800. 80*s.*

The striking characteristics of this large book are these: it is written by a middle-of-the-road psychologist. It adheres to the classical quadri-partite division into sensory, perceptual, learning and symbolic processes. It seeks to cover the whole of this field. It has a bibliography of some twelve hundred titles.

The author's aim has been to provide 'a text that evaluates experimental literature in close relation to critical theoretical issues' giving 'a more detailed description of critical experiments' than is usual, and attempting 'to provide an interpretative framework that both brings out the significance of existing research and directs the reader's thinking toward as yet unresolved problems' rather than 'simply laying out a series of separate studies'.

Within the traditional framework adopted, the range and coverage in detail of the experimental work that is discussed leave little room for criticism. There is a wealth of excellent graphs and diagrams. In some places there are perhaps fewer tables than one might like to see. Partly on this account, the danger is if anything that the student who seeks his information here may be lured away from the 'experimental literature' altogether, with adverse effect on his own powers of appraisal and criticism of original data. The author, too, largely fulfils his claim that he has tried to write so as to indicate unsolved problems to the reader. Necessarily his suggestions are based upon his own conceptions; to this extent he must at once arouse reservations and provoke controversy, if not outright disagreement; but there can be no doubt that this form of presentation makes for a more stimulating text-book than would any mere classified directory of experimental psychology.

It is when one looks for 'an interpretative framework' in a book about 'method and theory in experimental psychology' that one finds less than might have been expected. The book does not attempt to approach the whole of experimental psychology consistently from one theoretical standpoint. Though foreshadowed in a discussion of constancy in perception, it is not until half-way through the book that the author's main theoretical contribution is encountered; at the end of a chapter devoted to theories of learning he discusses his 'mediation hypothesis' in attempted reconciliation of the field theoretical and behaviouristic approaches to learning. It would be inappropriate to embark upon a detailed criticism, but some comments are in order in examining the approach to theory and theories that the book adopts. The hypothesis is not put forward following any methodological analysis of other theories. It can be held that there is a failure to appreciate radical differences between them at this level and, to this extent, that the mediation hypothesis provides an eclectic basis for the psychologist whose object is to carry out experiments, rather than that it affords any substantial theoretical reconciliation. Much the same commentary is to be passed upon the partial discussions of other problems, met with here and there through the book, such as that of similarity and generalization, and those that arise in relating physiological, behavioural and phenomenal evidence.

In reply the author could justly observe that he would never have written a text-book at all if he had set himself a different kind of task, and could point to what he has done. What he has done is indicated by such a chapter heading as 'certain controversial issues in learning theory'. A careful statement of the different current positions on particular issues is given, selected experimental evidence is presented and evaluated, and places where there is a lack of evidence are indicated. There are similar discussions of, for example, theories of thermal sensitivity and of the positions of Hebb and Gestalt psychology on perceptual organization. The level of theoretical evaluation is perhaps brought out as clearly as anywhere in the chapter on 'serial and transfer phenomena'. Here the author develops, as he has less fully elsewhere, a systematic approach to much seemingly contradictory evidence, the discussion culminating in a theory at the level of empirical generalization of a high order. Crucial to the theory is, however, a quantitative conception of *similarity*. The issue of whether, and if so how, a theory can usefully employ such a concept is not pursued, even though in another chapter the possibilities of circular reasoning that inhere in its use have been briefly indicated.

'Methods and theories in experimental psychology' might therefore have been a more appropriate title. But again, these terms are apt to be all things to all men and perhaps particularly to psychologists. It remains to welcome a refreshing text-book which is from its standpoint a stimulating account of a large part of what is current in experimental psychology.

ARTHUR SUMMERFIELD

Psychometric Methods (a new second edition). By J. P. GUILFORD. New York: McGraw Hill Publications Inc. 1954. Pp. ix + 597. 61*s*.

Many readers will be familiar with the first edition of *Psychometric Methods* which appeared in 1936. In the past, students of psychology have often turned to Guilford for statistical guidance. The new edition of *Psychometric Methods* is both longer and more selective than the earlier one. The comprehensive treatment of such statistical topics as frequency distributions, the normal curve, correlation and regression has been omitted since these have been included in the second edition of Guilford's *Fundamental Statistics in Psychology and Education*, published in 1950. This step has enabled the treatment of psychophysical methods to be extended and comparatively recent developments as well as the traditional methods are included. A mere catalogue of the contents of the sixteen chapters would convey little indication of the scope of the present volume. However, a broad summary of the contents will be useful. The first three chapters are essentially introductory, ranging from an elementary discussion of psychological measurement and psychophysical theory to a mathematical introduction intended to supply the minimum of necessary background knowledge for the benefit of those without formal training in mathematics. The next three chapters (4–6) deal, in turn, with the classical psychophysical methods. Chapters 7–10 deal with progressively more recent developments in psychophysical methods ranging from the Method of Pair Comparisons to various proposals for 'scaling' data based on human judgement. A lengthy chapter (11) is devoted to the Rating Scale as a psychological instrument. Whereas this forms an excellent summary of available facts, little seems to be added to what was already known and compiled. A most important chapter (12) is devoted to Principles of Judgement. In this Guilford summarizes the main factors influencing human judgement, and in doing so collects together for the first time the results of many recent psychophysical studies. Some of these were first published as late as 1952. From the point of view of the interested specialist this chapter will be one of the high-lights of the present volume. Guilford now turns his attention to a different set of topics. The next three chapters (13–15) are devoted to various psychometric problems in the field of mental testing. Such questions as item analysis, homogeneity and heterogeneity of tests, speed versus power tests, reliability and validity are discussed. The chapter on Test Development includes a brief reference to bias in the response to test items and methods which may be employed to minimize such bias are outlined. Attitude-scale construction occupies a few pages at the end of this chapter. The last chapter presents an admirable exposition of the theory and practice of Factor Analysis with emphasis on the American point of view. It is no mean accomplishment to have stated so much with clarity in so short a space. One may not agree with current American practice in Factor Analysis, but since it represents one important thread in contemporary psychology it is desirable for a concise yet effective statement of the procedures in use to be available. This chapter provides such a review, including worked examples.

The typography and layout of this volume are excellent, and remarkably few errors have been detected. Each chapter includes a lengthy bibliography and a few problems. At the end of the volume are appendices A–M which consist of a variety of statistical and mathematical tables in common use.

A critical review of *Psychometric Methods* would require more space than is available here, and more time for consideration than can be devoted to this review. However, a rapid appraisal is desirable. The student who wishes to obtain guidance on how to carry out one of the many specific techniques dealt with in this volume may consult it to his advantage. However, the more sophisticated reader will often retain mental reservations concerning the assumptions which have to be made in order to proceed with the more elaborate techniques. In a volume this size more space might profitably have been devoted to the qualitative implications of the assumptions which particular techniques force upon the investigator, and of the precise meaning of their end-products. It is quite clear from short interpolated paragraphs that Prof. Guilford is well aware of the necessity of stating clearly what assumptions can legitimately be made. However, he repeatedly throws upon the reader the responsibility for a decision on this point in applying the techniques outlined to practical problems. The beginner, for whom the mathematical introduction is intended, will not be in a position to decide such questions for himself. Furthermore, the sections on psychological scaling, on the psychometrics of mental testing and on Factor Analysis seem to be more advanced than the non-mathematical reader can approach in an intelligent and critical manner. Readers in the United Kingdom may regret the scanty reference to relevant work originating in this country. However, much of this work could not have been taken into account without reducing the reference to American work which is perhaps less accessible to many readers.

It might well be argued that we have here not one but the makings of more than one volume. Although one might just agree on the relevance of Factor Analysis to the main theme the three chapters on psychometric problems of test development seem to belong to a different 'dimension' from the rest of the book. Individual sections will be of value to the general reader, and will certainly stimulate the expert in the field concerned. *Psychometric Methods* is probably best regarded as a reference book which should be included on the shelves of the psychological section of all libraries. Indeed the present reviewer had sufficient confidence in the importance of what Prof. Guilford writes to recommend that this new edition be purchased without having first scrutinized it in detail. This action has been fully justified by the book itself, even if one does not share all opinions expressed or the assumptions made.

ROLAND HARPER

Motivation and Personality. By A. H. MASLOW. New York: Harper. 1954. Pp. xiv + 411. $4.50.

We have here what is really a collection of essays, most of which have appeared in earlier versions as journal articles. The author champions his socially conscientious, 'holistic-dynamic' conception of science in general and psychology in particular, and then proceeds to outline his well-known hierarchical theory of motivation. The theory distinguishes five levels to which the various motives of human beings can be assigned: physiological needs, safety needs, belongingness and love needs, esteem needs, and needs for self-actualization. The needs belonging to each of these levels emerge when those of the preceding level have been gratified. Maslow identifies progress along this sequence with an increasing degree of mental health. This leads him to recommend gratification of needs as far as possible and to feel that satisfaction of thwarted motives plays a greater part in psychotherapy than is usually acknowledged. The later chapters develop the implications of this theory for psychological health and ill-health, psychotherapy and cognitive processes.

Very little of the book is concerned with specifying the predictions that would enable one to test the theory or to see exactly how it differs from alternative theories. The evidence is mostly anecdotal, including case-history-anecdotal, or derived from clinical impressions. There are cursory appeals to controlled studies whose precise implications would bear prolonged discussion. Some interesting passages are sandwiched between generous slices of persuasive language, exhorting the reader to change to the author's preferred brands of classification and imagery.

Maslow's 'holistic-dynamic' principles seem to be adequately summarized by the assertion that behaviour is strongly influenced by motivational factors, and the assertion that behaviour is a non-linear function of a large number of variables. One would have to go back a long way to find a reputable psychologist who was unaware of these truths. But Maslow feels impelled to spend an astonishing amount of time berating a troop of straw men on dead horses, who, one gathers, make up the bulk of contemporary psychologists. These benighted individuals adhere to an 'atomistic-mechanistic-static' point of view: they appear to think that every response has a single external cause and is quite isolated from every other response, to explain learning by fortuitous combinations of external events which have no relevance to biological needs, and to have a liking for animal experiments without having realized that subhuman and human animals are different. He admits in the preface that these polemics, originally written about 1940, are now 'somewhat dated'. One would have thought that even such pioneers as Pavlov and Thorndike had written enough to absolve them from these errors, while their various successors have been proclaiming their avoidance of them for a quarter of a century. It is noteworthy that Maslow, who at other times provides copious references, gives none at all when attacking 'behaviourism' and that the abundant works of neo-behaviourists on the topics he handles are all but completely absent from his bibliography.

Maslow begins the book with a protest against an 'excluding, critical' view of science and in favour of a 'creative and constructive' one. It has evidently not struck him that people do not have to wait for scientists to supply them with ideas, least of all about human beings. But what has always been badly needed is a means of distinguishing the true from the false among the numerous hypotheses that are on the market. Some readers will fail to see why a strict application of the canons of scientific inquiry should be the same as a narrow one. To them the word 'science' will be inseparable from the ideals of precision and rigour and from a humble and critical attitude even to the assumptions one finds most acceptable. They will prefer unsolicited testimonials from clear-cut empirical data to the competitive skills of the copywriter, and the history of other sciences will convince them that these tastes are not inimical either to creative imagination or to social significance. Such readers will find the role assumed by Maslow too near that of the sage or the publicist for their liking.

D. E. BERLYNE

Learning Theory, Personality Theory and Clinical Research. The Kentucky Symposium. London: Chapman and Hall. 1954. Pp. 164. 28*s.*

Psychological literature has, of late, been increasingly encumbered by the publication of symposia in book form. More often than not, these lose in coherence what they gain in richness, the final product being a jumble of miscellaneous discourses rather than a unified treatise. So with this volume. It comprises eleven lectures delivered at a symposium designed to encourage a clearer integration of the three branches of psychology named in the title. The lecturers were: D. K. Adams, R. B. Ammons, J. M. Butler, R. B. Cattell, H. F. Harlow, N. R. F. Maier, O. H. Mowrer, D. Snygg, K. W. Spence, D. D. Wickens and J. R. Wittenborn. Taken individually, the lectures are able, stimulating, and informative. They reflect the characteristic research interests of their authors and give little cause for complaint. Collectively, however, they are less satisfying. There is but slight inter-relation between them and the co-ordinating pencil of an editor is nowhere evident. It is perplexing to speculate on the type of reader at which this work is aimed. For the layman, the lectures are too specialized. For the research worker, they are insufficiently detailed. And the professional psychologist can turn to recent issues of the *Psychological Review* and find there a collection of papers very similar to and as diverse as those published here. It is never unrewarding to read what such distinguished writers as the above have to say. But this fact, in itself, seems to lend slender justification to the production of a book which cannot be said to fulfil the ambitious and laudable intention of those who organized the symposium.

I. M. L. HUNTER

Studies in Schizophrenia: a multidisciplinary Approach to Mind-Brain Relationships. Reported by ROBERT G. HEATH. Harvard University Press; London: Geoffrey Cumberlege. 1954. Pp. xiv + 619. $8.50.

Schizophrenia has been investigated by many different techniques in many disciplines. Thus, at the beginning of this monograph Dr Heath, the chairman of this multidisciplined group at Tulane University, has at his disposal a number of ideas from psychiatry, psychology, physiology, biochemistry, neurology and neurosurgery, Under the title 'Theoretical Concepts', a unified approach is sought to provide a framework for the study for schizophrenia. Despite the fact that the work starts off with a discussion of theory, the empirical nature of the technique eventually devised cannot be disguised. This by itself is not criticism of the methods used in studying the disease of schizophrenia, or the therapeutic approaches suggested by these methods. The empirical is the contemporary rule in these matters. However, a more formal and critical review would appear necessary before the claims of the subtitle of the work could be upheld.

Briefly the formulation is that patterns of overactivity and underactivity, of motor, psychological and chemical homeostatic mechanisms are attributable to excitation of hypothetical circuits in the brain. These facilitatory and inhibitory circuits exist as cortical to subcortical connexions.

From electrophysiological experimentation and animal investigations, attention has been drawn to the importance of the dynamic interaction between cortex and the septal region of the forebrain. In schizophrenia, it is argued, an impairment of facilitatory excitation is involved. Therefore stimulation of the relevant circuit at the level of the septal region will produce beneficial changes in those patients suffering from schizophrenia. This is tried on a group of twenty patients, mainly chronic schizophrenics.

The technique of stimulation is by surgical or stereotaxic implantation of electrodes in the relevant subcortical structures and by bringing the leads out to a plug attached to the skull. Thus patients can be treated by subcortical electrical stimulation at some time after the neurosurgical procedures. Some degree of therapeutic success is noted, but fortunately the monograph is content to point out therapeutic possibilities only and not to advance therapeutic claims.

All this, of course, is by and for electrophysiologists, neurosurgeons and the like. The course of the investigation is dictated by their interests and techniques. The psychologist and psychiatrist play a secondary role in trying to describe modes of changes and directions of clinical improvement. In this respect the psychologists involved have profited by their experience in the Columbia–Greystones project and have contented themselves with a narrower range of investigation—namely to find a battery of tests which reflects the changes produced in patients by this stimulation technique. The battery in question included the Eisenson Tests of aphasia, a shortened Wechsler-Bellevue, Porteus Mazes, Benton Visual Retention Rest, Rorschach, T.A.T., a Behaviour Rating Scale, and a group of psychomotor tests which seem highly appropriate to this kind of work, but which disappointingly are not discussed or described.

Dr H. E. King, their author, however, apparently intends shortly to communicate a description of these tests and their uses.

While the psychological contribution by itself does not add much to our little knowledge in this field the monograph as a whole represents one extension of present physical methods of treatment which may in the future concern clinical psychologists.

R. M. MOWBRAY

Fundamentals of Psychoanalytic Technique. By T. BRAATØY. London: Chapman and Hall. 1954. Pp. ix+404. 48*s*.

In this book the late Trygve Braatøy gives an account of the methods of analytical psychotherapy which he developed as a result of a very large psychiatric experience in his native Norway, in France, Germany and the United States.

Braatøy was much influenced by the developments of Freud's ideas by Otto Fenichel and Wilhelm Reich before the second World War. Because of his wide interests in clinical medicine, physical medicine and medical ethics as well as psychiatry he became increasingly worried by the tendency of psychoanalysis to become divorced from other forms of medical treatment. Gradually he evolved his own principles and methods of analysis and in this book he describes them.

Because of the demands analytical treatment makes on the energy and time of the patient and the therapist and because it may be dangerous psychoanalysis is restricted to a relatively small number of patients who should be carefully selected. The author gives his criteria for deciding that analysis is indicated and stresses that account must be taken of the qualities of the analyst as well as those of the patient.

The major part of the book consists of detailed descriptions of the author's principles and methods of treatment, the difficulties encountered and the means to avoid them. He presents numerous case-histories and it is refreshing and very instructive that he discusses his failures, as well as his successes.

The book is practical in its approach. The position of the patient on the couch and the need for him to be relaxed are discussed in much greater detail than in Freud's concept of personality structure; techniques, not ideals, are its concern. At the same time the author justifies his methods with reference to a wide range of analytical and non-analytical literature. His common sense and insistence that accurate observation and diagnosis are essential for satisfactory psychotherapy leads to some penetrating criticism of a number of latter-day Freudians. But his criticism is of their practice not of their principles, and it is constructive, not destructive. In particular his review of the methods of Reich and Fenichel is most illuminating. The author's concern with technique rather than theory is illustrated by the fact that there is no reference to the works of Jung in the bibliography. It might well have been called 'Psychoanalysis in practice; a personal interpretation of Freudian principles'.

Braatøy wrote enthusiastically and the book tends to be diffuse and repetitious in parts. The English is usually adequate to make clear the meaning of the author but is never elegant. Norwegianisms abound. Many of the very interesting footnotes could easily have been incorporated in the text. As it is they are so numerous that they prove distracting. Since the book is written in English it would have been a convenience if translated German works had been quoted in their English versions in the bibliography.

The author states that he assumes his readers have a considerable familiarity with the major works of Freud. At least some experience of the practical problems of psychotherapy is required for the book to be fully appreciated. It is in no sense an elementary text-book and the student would probably be puzzled by it rather than enlightened. It is a fascinating and valuable account of the practical problems encountered by one courageous and individualistic psychiatrist in attempting to reconcile the principles and methods of psychoanalysis to those of other forms of psychiatric and medical treatment. The fact that the verdict must be that the reconciliation he accomplished remains incomplete makes one the sorrier that this is the last book we shall have from this stimulating psychiatrist.

T. T. S. INGRAM

The Collected Papers of Otto Fenichel. First Series. Collected and edited by HANNA FENICHEL and DAVID RAPAPORT. London: Routledge and Kegan Paul Ltd. 1954. Pp. xiv+408. 35*s*. net.

The Collected Papers of Otto Fenichel are being published in two volumes; and the present volume represents the First Series, consisting of papers written between 1922 and 1936. Fenichel, who died in 1946, was a psychoanalytical writer of unusual distinction; and his book *The Psychoanalytic Theory of the*

Neuroses is a classic work. An as expositor of Freudian theory he is unrivalled; and indeed his grasp of essentials and capacity for systematization, combined with his terse, lucid and forthright style, often render his expositions of classic psychoanalytical theory considerably more effective than those of Freud himself. At the same time, his sincere adherence to basic Freudian principles (and his obvious sincerity cannot fail to impress the reader) does not in any way impair the exercise of his critical faculty; and he does not hesitate to exercise this faculty, with a respectful cogency which sets a standard for scientific argument, whenever he feels criticism of any specific aspect of Freud's thought to be justified. Thus his criticism of Freud's theory of the death instinct in a paper devoted to this subject is perhaps the most cogent criticism of this theory on the basis of Freud's own principles that has ever appeared in print. The two papers entitled 'Psychoanalytic Method' and 'Concerning the Theory of Psychoanalytic Technique' contain masterly expositions of the principles underlying psychoanalytical therapy in its classic form. Two other important theoretical papers are those entitled 'Identification' and 'Defense [*sic*] against Anxiety, Particularly by Libidinization'. The remaining papers are predominantly clinical; but many contain theoretical formulations of considerable interest and importance. The depth of Fenichel's clinical insight cannot fail to command the respect of even the most experienced psychoanalyst. His elaborate interpretations of some of the clinical material which he describes may well mystify the reader with no special psychoanalytical training; but to the trained psychoanalyst they will present themselves as evidence of an unusual grasp of the complex currents of unconscious motivation on the part of Fenichel. The book as a whole is one which is bound to appeal to all who are interested in psychoanalysis, and will stir in all such the hope that there will be no undue delay in the appearance of the Second Series of Fenichel's papers.

W. RONALD D. FAIRBAIRN

The Development of Personality. By C. G. JUNG. London: Routledge and Kegan Paul Ltd. 1954. Pp. vii+235. 25*s.* net.

This is volume XVII of *The Collected Works of C. G. Jung*, and the fourth volume of the series to be published. It comprises eight papers, of which all except the last two ('The Development of the Personality' and 'Marriage as a Psychological Relationship') are devoted to child psychology and education, and are addressed primarily to educationists. According to Jung, the psychological development of the child consists essentially in the development of consciousness out of the original unconscious, which is partly collective and partly due to an 'identity' of the child's psychic state with the unconscious of his parents. Differentiation of the ego is regarded as characteristic of this process of development in so far as its aim is successfully achieved; and the first six papers are devoted largely to a description of the process in question. From a practical standpoint, the most important bearing of Jung's theme arises out of his view that the psychological problems of the child are chiefly a reflexion of the unconscious problems of his parents, and secondarily those of his educators as parent-figures. To the reader who has found light thrown upon the development of the child's ego by the views of Freud and Melanie Klein, doubtless the most remarkable features of these papers will be Jung's soft-pedalling of infantile sexuality and his minimizing of the significance of aggression; but the reader will also perhaps note with surprise the correspondence between Jung's theory of the development of consciousness out of the unconscious and Freud's theory of the development of the ego out of that part of the id which makes contact with outer reality.

The last paper, on 'Marriage as a Psychological Relationship', is devoted essentially to a favourite theme of Jung's later writings: viz. (as indicated in an Editorial Note) his conception of personality as 'an adult ideal whose conscious realization through individuation is the aim of human development in the second half of life'; and, in accordance with this conception, Jung expounds the view that it is only as the result of successful individuation achieved in middle life that marriage can acquire the status of a true personal relationship. The second last paper, on 'The Development of Personality', is essentially a description of that sense of vocation which has characterized the leading figures in human history; and for many readers this will doubtless prove the most interesting paper in the book. The scientific psychologist may be disappointed to find that Jung's illuminating description of what is involved in the sense of vocation is not matched by an explanation of its psychological origins; but he may also console himself with the reflexion that, when reductive explanations of such sublime phenomena are attempted (e.g. along Freudian lines), much of what is to be explained is commonly lost in the very process of explanation.

W. RONALD D. FAIRBAIRN

The Prediction of Performance in Clinical Psychology. By E. LOWELL KELLY and DONALD W. FISKE. Published by the University of Michigan Press, Ann Arbor. 1951. Pp. xv + 311.

This is an account of a large-scale research into the prediction of performance in clinical psychology. The authors tell us that at least ten Ph.D. theses went into this piece of work which stretched over some five years. The contributors to the project included many well-known names in the field of mental testing and personality assessment.

Almost every known method was used to predict aptitude for clinical psychology. From a study of this mass of material, the authors state that the Miller Analogies test was more significant than the Thurstone's P.M.A. and that less weight could be given to the M.M.P.I., the Allport-Vernon Scale of Values and to most of the Guilford-Martin Scale. Some sections of the Strong Interest Blank were useful. The user's confidence in the interview is an 'extremely fallible index of its actual validity in a specific situation'.

The various members of the research team were also equally thorough in their search for suitable criteria to take as measures of achievement in clinical psychology. Finally, they arrived at thirteen variables which include academic performance, clinical psychology and general psychology, diagnostic competence, individual therapy, group therapy, research, supervision, interpersonal relations, integrity, clinical competence, preference for hiring, and liking. The criteria were factor analysed and as a result three factors were identified which seem to characterize clinical performance. They were an *intellectual factor*, effectiveness in close *interpersonal relations* and *personal integrity*.

The authors tackle the problem of validity piecemeal after indicating the general pattern of the prediction problem (p. 124). First the validity of university predictions over a period of three years is considered. These are generally the highest of all the correlations obtained. Then the validity of the objective tests are set out and the intellectual ability tests are taken first. The Miller-Analogies test comes out very well with correlations of the order of 0·35 and 0·47 for various parts of the multiple criterion. The *R* and *M* subtests of the P.M.A. test are most suggestive of prediction and in the Guilford-Martin battery, thinking extraversion, freedom from cycloid tendencies, and social extraversion yield the highest correlations with related parts of the criterion. These include diagnostic competence, individual psychotherapy, interpersonal relations, integrity and preference for hiring. Research competence, academic performance, clinical competence, preference for hiring and the marks in clinical and general psychology are predicted well by the group I of the Strong vocations, particularly those referring to the artist, psychologist and architect. In the group II occupations, the mathematician and physicist responses correlate with academic performance and research competence. A review of the higher correlations obtained from these different objective tests is given on page 158. The validities of the personality assessment ratings are, on the whole, quite high, ranging up to 0·5. One interesting result shows that intervening personal therapy tends to reduce the validity of personality assessments.

It is a pity that the authors did not have recourse to multivariate analyses in order to obtain the maximum prediction from a selected number of predictors. If it were objected that such a task would be extremely lengthy, with so many components in the criterion and in the predicting side of the battery, one could suggest that use might be made of the discriminating function. The book is not easy to read, particularly for the English reader. It presents a trying mixture of heavy passages of technical jargon interspersed with tables of extremely condensed material. Lastly, it is, perhaps, somewhat chastening to find that certain tendencies already discovered in less ambitious programmes were merely confirmed.

However, the work as a whole should be studied by all engaged on the prediction of success in a complex professional field. The 'general outcomes' listed on page 197 are valuable and follow from the data. Some of the methods described and the results obtained are very pertinent to current studies being carried out in this country on the selection of medical students and teachers and for other professional occupations. E. A. PEEL

Deprived Children. The Mersham Experiment. A Social and Clinical Study. By HILDA LEWIS, M.D., M.R.C.P., with a foreword by Sir ARTHUR ELLIS and DR C. P. BLACKER. Published for the Nuffield Foundation. London: Oxford University Press. 1954. Pp. xiii + 163. 9*s*. 6*d*. net.

This book is a scientific assessment and evaluation of the work of the Mersham Reception Centre. This Centre arose as a pilot experiment on the instigation of Miss Leila Rendel, the Director of the Caldecott Community, to the Nuffield Foundation. Miss Rendel was 'gravely disturbed at the haphazard way in which deprived children needing care were distributed to children's homes, foster homes or approved schools without any real study of their individual needs or difficulties'. As a result the experiment, which was initiated in 1944, was sponsored by the Caldecott Community whose 'solid background of achievement and experience in the case of deprived children was available for this first study of the problems involved in a Reception Centre'.

A description of the Reception Centre and its work is given. A high proportion of the neurotic and maladjusted children were referred to the Centre. One-twelfth of the children received were recommended for return to their own homes, one-fifth to foster-homes and slightly more than one-third to children's homes. 76% of the children were placed according to the recommendation made. Certain practical difficulties arose because facilities in the county of Kent were restricted, although these became easier after the passing of the 1948 Children Act.

A fairly detailed account is given of the 500 Kent children admitted between October 1947 and July 1950, 277 (55%) of whom were boys. Information was obtained on their family backgrounds in terms of social class, size, income and living conditions, and a careful study was made of the 363 families, information being available concerning 264 fathers and 328 mothers. There is evidence to show that the families were larger than the average for the county, that their incomes and occupational stability were low and that a fair proportion (30%) were living in over-crowded conditions (comparable figure for the general population of Kent was 1·09%). Relationship between the parents was estimated as definitely bad in 45% of the families. Medical disabilities especially mental disabilities, both defects of intellect and personality, were common (99 fathers and 180 mothers).

A quarter of the children admitted to the Centre were normal in behaviour and general mental condition. 4% were classed as psychotic or psychopathic, 32% as delinquent in some degree and 18% showed severe neurotic symptoms. There was a slight preponderance of dull children, but 47% were of average intelligence. Attainment in reading and arithmetic was poorer than age or intelligence warranted. 21% were in poor physical condition. It is noteworthy that the two most marked psychological characteristics of the children were excessive demands for attention and lack of emotional warmth in their reactions.

An attempt was made to classify the general behaviour of the children according to the patterns described by Hewitt and Jenkins, viz. unsocialized aggression, socialized delinquency, over-inhibited neurosis and normal behaviour. This is of interest to the psychiatric reader, but it is doubtful whether such a classification is of great usefulness.

The analysis of the data obtained by so qualified a team of observers as psychiatrist, psychologist and psychiatric social worker is of great value. It is important to find evidence to show that five background features had a statistically significant association with patterns of disturbed behaviour, i.e. neurotic or psychopathic mother, a mother lacking in affection for the child, prolonged stay in public care and illegitimacy. It is valuable also to find the concept that early separation from the mother invariably produces serious emotional effects carefully examined. There was no evidence that delinquency or incapacity for affectionate relationships was significantly more frequent in the separated children. Children who had suffered lasting separation from their mothers, which began before the age of two, tended to be more disturbed in behaviour than those who had not been exposed to lasting separation at any age however. But the report emphasizes the importance of considering the quality of the child's relation to his mother before deprivation and his experiences with mother-substitutes while separated. It also stresses the emotional hazards and harms children may suffer from bad mothers and indifferent mother-substitutes and the need for more sober research into this whole matter.

There were 66 'problem families', characterized by gross neglect and squalor, represented in the sample. Although many of the parents were handicapped in personality or intelligence they were affectionate to their children, who though of lower intelligence, were freer from delinquency or neurosis than the rest of the Reception Centre children. This supports other findings in regard to this group.

The follow-up study which was made on 240 children after an interval of 2 years or more and included a special intensive inquiry on 100 children, revealed that the proportion of children in good psychological and social condition had more than doubled. Children who had been placed in accordance with the recommendations of the Centre had done somewhat better than those who had not been placed as recommended. On the whole, children placed in foster homes proved most satisfactory though not invariably so. Close contact with his own family proved greatly to the child's advantage in general. Good intelligence was also prognostically favourable.

Any serious reader of this book will be impressed with the care and scientific caution by which the work of the Reception Centre is evaluated. With what is described by Dr Blacker and Sir Arthur Ellis as the utmost emotional neutrality, Dr Lewis has shown the very great value of the work of a well-run, adequately staffed Reception Centre in assisting in the difficult and delicate task of finding the appropriate placement for the particular child. Her findings are supported by the Childrens' Officer of Kent. The various alternative methods are examined in the book and their disadvantages demonstrated and the experience and evidence accumulated at the Mersham Centre in Dr Lewis's words 'compel the firm conviction that reception centres, though not ideal, are indispensable at the present time for dealing with a large proportion of deprived children', but that 'a bad reception centre is probably worse than none at all'.

This is a book which all workers in the field of Child Care cannot neglect to read.

AGATHA H. BOWLEY

Children's Humor. By MARTHA WOLFENSTEIN. Glencoe, Illinois: The Free Press. 1954. Pp. 224. Price $3.75.

This book contains examples and analyses of children's humour at different ages. The analyses are based on the theory of Freud as developed in his *Wit and its Relation to the Unconscious.* In addition to its manifest content, the joke, like the dream, has also a latent meaning which has undergone distortion and disguise. In the introduction, the author suggests that some readers may be inclined to feel that more is seen in the jokes than is there, with which statement the reviewer tends to agree.

From an early age, the child uses the joke to enable him to alleviate difficulties. He endeavours to transform painful experiences into enjoyable ones. The basic motive is to triumph over distress and to gain relief, if even momentarily, from frustration. The 5-year olds try to cope, through joking, with their envy of the bigness and power of adults. From the age of 6–11, the latency period, jokes tend to be identified with riddles, and the 'moron' who represents all that the child repudiates in himself, becomes the main protagonist. At adolescence, the riddle is replaced by the anecdote, and mimicry becomes a component of joke telling.

The five chapters in turn deal with Joking and Anxiety; Sex, Names and Double Meanings; Riddles and the Legend of the Moron; Development of the Joke Façade; Understanding the Joke. The contents of the chapters are interesting, and the analyses of the jokes throw light on certain aspects of childrens' development, although one wearies of the constant Freudian interpretation.

The book can be recommended as a useful contribution to an understanding of humour, as well as a careful study of a specialized field.

M. COLLINS

Psychology of Learning and Teaching. By H. W. BERNARD. New York: McGraw-Hill Publications Inc. 1954. Pp. ix + 436. $5.00.

It is nowadays important to note how far the pupil is a product of social circumstance and the scope of educational psychology has widened accordingly. Dr Bernard thinks and writes in this tradition, stressing 'the shift in emphasis from physiological psychology to one oriented in social groupings'. So we have sections given over to 'the role of the social setting' and the teacher's 'dynamic influence' on the pupils' mental hygiene.

He underlines also the continuous nature of child development and 'growth as a pervasive factor in learning and teaching'. Finally, evaluation itself is seen as 'a continuous process rather than as a culminating act'. This is a view which will find sympathy in Britain, for we know only too well of the dangers of taking a single cross-section when we test progress and of the value of cumulative records.

The material in *Psychology of Learning and Teaching* is well presented, but is perhaps a little full and even wordy for a first text-book. Modern theories of learning, motivation and goal seeking are all invoked, but in spite of the useful summary provided at the end of each chapter, they are not sufficiently outlined for an inexpert reader. In general, the reader would be helped by a more explicit reference to theory. So on page 28 we read that 'not only are habits ways of acting, but, even perhaps more important they constitute a form of motive'. The reader's grasp of this important observation would have been enhanced had the author referred him explicitly to the concept of secondary reinforcement and related this to the other main features of Hull's theory. The main purpose of theory is to provide a comprehensive explanation and to give the teacher a frame of reference by which to relate and assemble the facts of his teaching experience.

The omission of these references to theory tends to give the book the appearance of a book of tips for learning and teaching. The British student might tend to find the scene and examples a little alien to him —as, for example, in the 'time schedule' and study tips given to student readers on pp. 26–27.

As a whole, however, the book is a useful description of many of the facts of learning, motivation and individual differences.

E. A. PEEL

A Theory of Social Control. By RICHARD T. LAPIERE. New York, Toronto and London: McGraw-Hill Publications Inc. 1954. Pp. xi+568. 53*s*. 6*d*.

This is a book for sociologists. Its thesis is that man is a 'status seeking animal' and that this status is sought in relatively small and intimate groups which thus provide the medium in and through which social norms are operated and social control of behaviour is brought about. Thus, for the author, social control is one of the three major determinants of human behaviour, the other two being personality and situation.

It will be apparent that such an analysis depends for its cogency on clear, full, and somewhat arbitrary definitions of the factors involved. For LaPiere the distinguishing feature of social control is that it operates through the conferring or withholding of status, and the major part of the book is concerned with the techniques by which this is accomplished in groups and societies of different kinds.

For the social psychologist, whose primary interest is, or should be, measurement and experiment in the social sciences, the second part of the book, on the nature and operation of social control, is the most interesting and the least satisfactory. The author's procedure of stating his hypotheses and illustrating them anecdotally and by reference to work which is, in most cases, of an anthropological nature, is here least effective. For example, we would agree, in a purely common-sense way with a proposition of such high generality as: 'All other factors remaining equal, the control that is exercised by a group over an individual member is inverse to the size of the group', although, if pressed, we would probably have little difficulty in finding exceptions. But when this statement is translated by the author into a proposition of lower generality, namely: 'Generally speaking, a group of ten can exercise ten times as much effective control over an individual as can a group of one hundred, a hundred times as much as a group of one thousand, etc.', then we feel that if words mean anything at all, it is time to call a halt and insist that rigorous, measurable definitions of the variables involved and experimental evidence of their relation, be forthcoming.

It should not be thought that the book contains many unsupported lower generality statements. Most of those presented are indeed the more sophisticated 'high' generalizations. This is unfortunate, since the use of 'high' language tends to obscure the necessity for empirical validating evidence; or to put it more bluntly, one can get away with more in a 'high' language than in a 'lower' one. (There comes to mind the paradox that Hull's greatness lies in the fact that he was the first psychologist who could be proved wrong.)

This work contains many interesting hypotheses for psychologists, but until the author uses a 'lower' language, we must regretfully repeat that this is a book for sociologists.

DEREK PUGH

Sex and the Social Order. By GEORGENE H. SEWARD. Harmondsworth: Penguin Books Ltd. 1954. Pp. ix+275. 2*s*. 6*d*.

This book was originally published in 1946. It must be said at once that it is a good book and fully justifies its appearance as a Pelican. It is exceptionally clearly written and easy to read, as well as informative, and can be recommended to anyone. The chapters are short and compact, and there are name and subject indexes and a bibliography of seven hundred items. If the author shows signs of suffragette zeal in her pleas for greater equality between the sexes, the book is none the worse for that.

Of the sixteen chapters, one is introductory and one, entitled 'Dynamics of Sexual Behaviour', deals briefly with the glandular and nervous basis of sexual behaviour. Five chapters are devoted to sexual behaviour in the animal world, from fish to chimpanzees, and four chapters to sex values and roles in primitive and in western society. Four chapters review the stages of sexual development from infancy to old age, and some of the associated problems, especially but not exclusively the social aspects thereof. One chapter gives a rather sketchy discussion of sex differences in modern (American) society. The last chapter, headed 'Sex in Post-war Society' contains a vigorous statement of the author's plea, which has already been made explicitly or implicitly a number of times, for a reformulation of sex roles in the interests of mental health and social efficiency. At times, the author is perhaps a little uncritical, and some topics which appear relevant have been omitted, for example, the problem of differential fertility and its possible effects.

One is left wishing that the author had dealt with her subject at much greater length. But a longer book would perhaps have proved less enjoyable reading. D. GRAHAM

Character and Social Structure. By HANS GERTH and C. WRIGHT MILLS. London: Routledge and Kegan Paul. 1954. Pp. xxv+490. 30*s*.

In the preface to this book, the authors state that 'our general purpose is to study the personalities of men in connexion with types of social-historical structure. We wish to analyze conduct and character by understanding the motivations of men who occupy different positions within various social structures'. This promises well, but to at least one reader, this promise is not fulfilled, and the book is to this extent disappointing. Despite the subtitle 'The Psychology of Social Institutions', the book is much more of an essay in sociology than a contribution to psychology. In Part One, the authors distinguish between organism, psychic structure ('integration of feeling, sensation and impulse'), person ('man as a player of social roles') and character structure ('integration of the organism's psychic structure with the social roles of the person'). One might expect most of the book to be concerned with character structure as thus defined.

Part Two, entitled 'Character Structure' draws heavily on the ideas of G. H. Mead. There is an interesting chapter on the sociology of motivation and the attribution of motives. What strikes the reviewer most, however, is the absence of any proper attempt to link the analysis of character with learning theory.

But most of the book is concerned with analysis at the institutional level. Parts Three and Four, entitled 'Social Structure' and 'Dynamics' respectively, are uneven. The chapters on 'Symbol Spheres' and 'Stratification and Institutional Orders' are rewarding, those on 'Institutions and Persons', 'Institutional Orders and Social Controls' and 'Collective Behaviour' are less enlightening. Material is drawn from historical periods as far apart as Ancient China and the contemporary world. The overall result is somewhat disjointed, and one wonders whether it would not have been better to restrict the canvas in one way or another. In particular, one may question the advisability of trying to deal, even in a cursory way, with the characteristics of the principal world religions in less than four pages.

If this book had to be summed up in a single sentence, it would perhaps not be unjust to say that it has attempted too much at the institutional level, and achieved too little at the psychological level. This is, of course, a psychologist's judgement. D. GRAHAM

The Psychology of Politics. By H. J. EYSENCK. London: Routledge and Kegan Paul. 1954. Pp. xvi+318. 25*s*.

When we come into contact with clear, original and forcefully expressed ideas, it is almost axiomatic that we say, perhaps somewhat wistfully, 'Now why didn't I think of that myself?' It sounds so easy, and Prof. Eysenck's conclusions have just this quality of deceptive simplicity. But what the man in the street tends to forget is that accurate simplicity, as opposed to naïve hit or miss, is hard work. The detailed experimental work and the tough conceptualizing which have enabled Prof. Eysenck to summarize his conclusions in a few lucid pages, compel respect and admiration.

This book, selective as science must be, does not pretend to describe all the historical, economic, and religious, etc., ideas which govern political affiliation. It is merely concerned to discover whether anything can be said from the psychological point of view; whether a psychological approach can reveal qualities held in common by persons of the same political outlook. Briefly, the conclusions are as follows.

Political attitudes are governed by two quite independent factors: the Radicalism–Conservative continuum (Prof. Eysenck refers to its as the R-factor); and the Tough–Tender minded dichotomy of William James (the T-factor). Of these two, the R-factor is social in origin, while the T-factor is personal. Thus fascists are tough-minded conservatives and communists tough-minded radicals. Liberals are tender-minded and in the middle of the Radicalism–Conservatism continuum. In addition, there is a relationship of tough-mindedness with extraversion, with aggression and dominance and with 'such concepts as rigidity, intolerance of ambiguity, narrow-mindedness and mental concreteness'. The tender-minded, by contrast, are introverts.

These formulations would not, of themselves, prove of essential value to psychology, however much light they might throw upon the social scene, if they were not linked with more general issues. In particular, Prof. Eysenck attempts to relate his findings to learning theory since attitudes, being acquired dispositions, closely resemble habits. Viewed in this context, R and T-factors are respectively the neural modifications involved in learning and in conditioning, the former implying a change of behaviour through the influence of reward and punishment (and hence social in origin); the latter implying changes of behaviour derived from the involuntary emotional responses evoked by association.

This is the meat of the book. But for the specialist, its value will also lie in the detailed account of experimental method and procedure, in the honest exposition of uncertainty, and in the commentaries made *en passant* on much other psychological work.

This admirable book may not make the pioneer impact of *The Authoritarian Personality* by Adorno and his colleagues, but there was an undertow of emotion in that remarkable work which aroused certain doubts. In some respects, Prof. Eysenck has made a more solid contribution to psychological theory. In particular, and this is always the test, he has opened up many new fields for future research.

ADAM CURLE

The University Teaching of Social Sciences: Sociology, Social Psychology and Anthropology. UNESCO. 1954. Pp. 525. 10*s.* 6*d.*

This is the first of five UNESCO reports comparing the teaching of the Social Sciences in different parts of the world. The International Sociological Association carried out a survey of the extent and organization of teaching and research in eight countries. A 'rapporteur' was appointed for each country (Donald MacRae here), and the reports were discussed by a committee. Prof. de Bie of Louvain gives an overall review of the results, and there are several shorter articles, including one by Prof. Nuttin on Social Psychology, as well as factual summaries for the different countries covered. This arrangement involves a lot of repetition and could be improved upon in the later reports.

It is no surprise to read that there is a tremendous gulf between America and other countries in the extent of social psychology teaching and research. Furthermore, the subject has a broader scope there, while there is greater development of standardized techniques of research, and more team-work. From what information is given the subject has much the same content in many other countries as it has here, and the same American books are used, save in Germany and parts of France, where the old philosophical approach may be found.

This is a useful little book, though the British material is not always accurate—partly through being four years old now. A number of recommendations are made—that teaching should be linked with research, that better training in research methods be given, and that close co-operation with experimental psychologists and sociologists should be maintained.

MICHAEL ARGYLE

The Lowenfeld Mosaic Test. By MARGARET LOWENFELD. London: Newman Neame Ltd. 1954. Pp. 300 + 144 colour plates (in separate container). 50*s.*

The Mosaic Test was devised by Dr Lowenfeld more than twenty years ago, but this is the first text-book which has appeared from her hand describing the test in detail.

The book is divided into two main sections. Part I contains a description of the test itself, and discusses the method of classification of the results. The test consists of 456 pieces composed of different geometrical shapes—squares, half-squares, diamonds, equilateral triangles and scalenes—each coloured in one of six colours, red, blue, yellow, black, green and white. The individual is asked to make anything he likes on the tray provided, and also given instructions to use as many or as few pieces as he wishes, and any of the shapes or colours. The shapes and colours used are significant, as well as the type of design produced, and its position on the tray. Dr Lowenfeld describes in detail the different types of designs, and their classification which is exceedingly complex. Constant reference is made to the 144 coloured reproductions of designs which accompany the book.

Part II discusses the use of the test with different groups, and illustrates how the test is interpreted. A separate chapter is devoted to each group: Children, Individuals with subnormal intelligence, The Normal Personality, Neurosis, Mental Disorders (written by Dr Ellenberger), Cultural Problems.

The interpretation of any design depends on six aspects: (*a*) the manipulation of the material; (*b*) the relation between the pieces used and the space of the tray; (*c*) the form of the resulting design; (*d*) the handling of colour; (*e*) the content; (*f*) the comments of the subject and his attitude to the test, and to what he has made. The author claims that the form of the Design represents a kind of self-portrait.

Some of the interpretations are extraordinarily interesting such as those of two sisters whose designs show striking similarity. The chapter on children's designs, tabulating the changes in the form of the design with age, is of value from the point of view of children's development. The designs indicating neurosis and mental disorders show features which differentiate them from the normal. Dr Ellenberger, who contributes the chapter on the use of the L.M.T. in the study of mental disorders, confirms Wertham's conclusions that the L.M.T. is an excellent and invaluable instrument for use in the differential diagnosis of diseases of the brain from mental diseases. The test, according to the same writer, has also yielded significant results with schizophrenics and can be used when other tests have failed. Dr Lowenfeld herself claims that the designs produced may detect nervous breakdown a considerable time before it actually occurs.

The different facets of the test seem unlimited, and if one can judge from the contents of the book, the test is being extensively tried out in different fields, and used in different countries. One considerable disadvantage from which it suffers is the difficulty of making a permanent record of the results. Apart from colour photography, white paper has to be placed on the tray beforehand, a tracing made before the pieces are removed, and finally the design has to be coloured—a time-consuming process.

A second disadvantage of the test, as it appears to the reviewer, is that the interpretation of the designs still seems to be largely subjective. As is stated in the book, to classify and to interpret the designs is like learning a new language, and the only reliable method of learning is said to be personal experience and direct personal teaching. 'It must also be granted that for the full use of the test and an appreciation of its subtler aspects, a certain special quality is needed in the investigator, one akin to the quality in the musician that makes elements of form and rhythm obvious to him that, to the unmusical, are imperceptible' (p. 48). If this is so, then the use of the test is certainly limited, and this unfortunately is the impression one receives from persual of the text.

While the book makes interesting reading, one would have liked the contents of some of the chapters to have been arranged more systematically.

M. COLLINS

The Thematic Apperception Test and the Children's Apperception Test in clinical use. By LEOPOLD BELLAK. New York: Grune and Stratton. 1954. Pp. x+282. $6.75.

As Dr Bellak points out (p. 147) TAT was originally developed by Henry Murray in an academic setting; not only the material but the general approach bears the mark of this origin, and it has been Bellak's aim to develop a technique of analysis and interpretation rather more fully appropriate to clinical use. Repeated reference is made to the need to accumulate norms, including separate study of different regional and cultural subsamples, but the author is careful to add that such norms are needed 'not to substitute for dynamic diagnosis, but to serve as a reliable background for it'. Bellak's major contribution in the earlier (and slightly longer) portion of the book, which deals with TAT proper, is the presentation of his own system of analysis, which, in so far as it is based on the use of a number of 'scoring categories' concerned more with content than with form, most closely resembles Wyatt's system (1947, *J. Psychol.* XXIV, 319–30), and stands in contrast to Rapaport's and to some extent to Murray's own, although something rather like need-press analysis is to be found in Bellak's listing of the 'main needs of the hero', together with 'conception of environment', 'parental (and other) figures', 'significant conflicts', 'nature of anxieties', etc. All this accurately reflects the author's clinical purpose; a more rigorous methodological innovation, reminiscent of Tomkins's approach, is his recommendation that the 'Main Theme' be progressively restated on various levels (up to five) of generalization or abstraction. Scrutiny of the data thus assembled, it is claimed, usually enables one to recognize a repetitive pattern in the subject's responses, and to proceed thence to a final report, which may or may not offer a diagnosis. Since he does not regard TAT as primarily a diagnostic instrument, Bellak recommends the use of some such formula as 'The data represented in the TAT are consistent with a diagnosis of...'. Such diagnosis, he says, should never be made on test evidence alone, without additional information provided by a clinical interview.

The second part of the book is devoted to a description of the Children's Apperception Test (CAT) and a Supplement to it (CAT–S). Based on a suggestion by Ernst Kris, the material consists essentially of animals mostly (though not always) in anthropomorphic guise. The CAT and the Supplement each consist of ten pictures, nearly all depicting or including 'animal children'. The situations in the CAT proper are regarded as more 'universal'; those in the CAT–S relate to more specific themes, such as temporary disability. In addition, the CAT–S pictures are better suited to more informal administration; they are on curvilinear cut-out shapes which ask to be handled. Most users of TAT proper will agree that even Murray's 'children's' cards are not entirely satisfactory for use with children, and Bellak makes out a fairly convincing case for the substitution of animal for human figures. The age range recommended is three to ten, and the material is claimed to be relatively culture-free. This latter seems to this reviewer to be rather a questionable assumption: not only the humanized animals but some of the family-life situations are unmistakably American. Accordingly, it comes as rather a surprise to read in the Appendix on 'Publications and Work in Progress', that extensive work on the CAT is going on in India, France, the French colonies and elsewhere. One also gets the impression that too close a preoccupation with some of the stock 'psychoanalytic' and allied themes (e.g. primal scene, castration fantasy, the body-image, etc.) has been allowed to divert attention from the possibly more superficial but nevertheless pressing problems which form the staple of child guidance work, at least in this country.

It has been necessary in this review to concentrate on the descriptive side of Bellak's work, but the book also contains a good deal of valuable and thought-provoking general discussion of projective theory and practice. It is eminently a readable book, the author's style being both succinct and at times pleasantly colloquial.

BORIS SEMEONOFF

Sentence Completion. By J. Q. HOLSOPPLE and FLORENCE R. MIALE. Springfield, Ill. U.S.A.: Charles C. Thomas; Oxford, England: Blackwood Scientific Publications Ltd. 1954. Pp. xiii + 177. 40*s*.

Subtitled 'A Projective method for the study of personality' this book presents a collection of 73 'openings', from which, it is claimed, the clinician may hope 'to obtain material from which [to] draw valid inferences concerning unconscious and semiconscious desires, motives, conflicts and systems of personality organization'. How this may be done is described almost exclusively by reference to illustrative records, which occupy about three-quarters of the book. Only one case, however, is interpreted at all fully; this subject's Rorschach record is appended, verbatim, but entirely without comment, except to call attention to 'the clear consistency which may be found when sentence completion and Rorschach materials are compared'. Lists of 'characteristic completions' are provided, but again without comment, and there is no statement of rationale, no mention of any definable technique of interpretation other than intuitive empathy, nor any data relating to standardization or to validity, which is assumed. Only passing reference is made to earlier work (e.g. Tendler, Rhode) in the same field. The book is attractively produced, but 40*s*. seems rather much to pay for so fragmentary a presentation of what is not, after all, an original idea.

BORIS SEMEONOFF

Personality in Handwriting. By ERIC SINGER. London: Gerald Duckworth and Co. Ltd. 1954. Pp. 120. 10*s*. 6*d*.

Graphologie et Physiologie de l'Ecriture. By H. CALLEWAERT. Louvain: E. Nauwelaarts. 1954. Pp. 168 + xxxi. 90 Belgian francs.

These two books on graphology offer a remarkable contrast. Mr Singer addresses himself to the wide audience of 'everybody who, up to now, has encouraged my efforts', and expresses the wish that his writings 'will also help to convince wider circles and will bring nearer the day of general official and universal recognition of graphology in this country'. Dr Callewaert, writing for a more specialist public, holds that graphology, understood as a means of judging personality from handwriting, is unscientific and based on no sound statistical or physiological data; the handwriting is not like the expressive gestures and mimicries that accompany speech; and that, like speech transmitted by microphone and so robbed of facial and gestural accompaniments, writing consists largely of laboriously learned and therefore conventional elements.

To psychologists who value the good name of projective methods in the assessment of personality, and who are wary of the 'blind-diagnosis' and the 'dream-book' kind of interpretation, the cold wind of Dr Callewaert's criticism is healthier and more acceptable than Mr Singer's partisan belief that, 'The graphologist can *seldom afford the luxury of interviewing his subject*. It is one of the most significant differences of graphology compared with other personality tests and projection methods that the graphologist does not need the presence of the tested person, and can work without questioning and interviewing' (p. 55, reviewer's italics). The approach to the subject is dogmatic and unscientific, and the book should appeal only to the uninquiring layman.

Dr Callewaert includes a useful historical summary on writing and graphology and devotes a large part of the book to a consideration of the physiological mechanisms involved in writing. Both books are freely illustrated.

LORNA M. SIMPSON

Die Tests in der klinischen Psychologie (*Tests in Clinical Psychology*). By E. Stern (ed.). Zurich: Rascher Verlag. 1954. Pp. 418. S.Fr. 28.10.

This is the first of two volumes dealing with 'tests in clinical psychology', to which authors of six different countries have contributed. The collection contains twenty-one papers, ten coming from France, five from Switzerland, three from the U.S.A., two from Germany, and one from Portugal. In 20 pages of introduction respectful homage is paid to about thirty scientists predominantly of the medical profession and of little or no historical interest to the problem. In no instance, however, is there reference to any living non-medical experimental psychologist working in the clinical or personality fields.

The book is divided into three parts. The first part contains the introduction, a description of the unstructured interview situation, as well as a historical survey of test development. The second chapter describes intelligence testing, developmental and motor tests, and some elementary statistical methods. Part three introduces the following tests: M.M.P.I. and some other questionnaires, Jung's association test, The Szondi test, Mira's drawing test, and two Swiss tests in which the subject's choice of book titles and proverbs is interpreted.

Most of the contributions represent introductory descriptions of the techniques, and contain little detailed discussion of relevant literature. There is, with the noteworthy exception of the American authors Hathaway and Meehl who deal with the M.M.P.I., little or no factual discussion of test construction, validity, or other statistical requirements. Further drawbacks are the failure to present more objective personality tests of the experimental type, the fact that the bearing of clinical methods on personality theory is not discussed, and that no indication is given of future trends of development.

It is hoped by the reviewer that the few trends towards more rigorous standards found in this publication will not pass unnoticed by the reader to whom it is addressed. The foreign reader can obtain from it a representative account of current continental ideas on clinical testing.

J. BRENGELMANN

Factorial Analysis for Non-mathematicians. By C. J. Adcock. Melbourne: Melbourne University Press. London: Cambridge University Press. 1954. Pp. 88. 17*s*. 6*d*.

The aim of this book is 'to cover the essential processes of factor analysis in such a way as to make clear the underlying logic in simple language'. The approach is through well-chosen numerical examples. Attention is focused on three methods: analysis for simple structure (Thurstone), cluster-directed analysis, and group factors derived from preliminary simple summation (Burt). Criticisms from the point of view of factor analysis are that the basis of factorial methods is not said to be measurements but correlations, that no account is given of how product-moment correlation comes logically to be involved, that use of the tetrachoric coefficient is even made to appear better, and that the mistaken procedure of inserting new communalities for the calculation of each successive centroid factor is advocated. The statistician can also point to the isolation of the discussion from any general consideration of statistical methods and their requirements. The main commentary is contained in chapters on the logic of factor analysis, rotation, interpretation of factors and the role of factor analysis. Here the arguments advanced can be held to suffer from their brevity and may be found persuasive rather than logically compelling. With qualifications, therefore, the book has much to recommend it as a teaching device, and it is of interest for its straightforward presentation and comparisons of the principal methods of group factor analysis.

ARTHUR SUMMERFIELD

Laboratory Instrumentation in Psychology. By W. W. Gringes. Palo Alto, California: The National Press. 1954. Pp. iii+282. $4.75.

This volume is described as an introduction to the laboratory for graduate students and for professional people who require sufficient detail to enable them to start work in a selected research field. Although the approach is necessarily selective, apparatus of interest to a considerable range of psychologists is considered. A number of the items described belong to the proverbial 'brass instrument psychology', but their modern electronic counterparts are also included. Each chapter is devoted to a separate field of instrumentation. Special attention should be drawn to the chapters on Audition, Vision and Bioelectricity. The first includes a useful discussion of the physics of sound and also deals in an elementary manner with sound-recording. The second includes a valuable survey of the nature and measurement of light and colour. The third starts with a survey of the types of electrical activity of special interest to the psychologist. The technical problems of electrodes and conducting media suitable for use in examining various forms of electrical activity in the human body are considered. This final chapter ends with an outline of the terminology and basic problems involved in amplifying such electrical phenomena. Each chapter is copiously illustrated with diagrams, photographs and actual electrical circuits where these are relevant. *Laboratory Instrumentation in Psychology* fills an important gap in existing psychological literature.

ROLAND HARPER

Minds and Machines. By W. SLUCKIN. Harmondsworth: Penguin Books Ltd. 1954. Pp. 223. 2*s*. net.

Mr Sluckin provides, what has not previously existed, a sound, straightforward and accurate account of the body of recent publication and discussion that passes broadly under the heading 'Cybernetics', particularly in its bearing upon psychology. He sustains a high and even level of competence and caution throughout—there is very little that is misleading in this book. Neither is there any false titillation of appetites to which some popularizations of Cybernetics—in common with 'Science fiction'—seem designed to pander. Equally there is little intrepid groping at fundamentals. The result is in many ways a satisfactory, and certainly a useful book.

The principles underlying the representation and communication of information, of mechanical computation, of negative feed-back and of servo-mechanisms are described in terms which are clear and easy to follow. Possible analogies between existing machines and the central nervous system are developed along familiar lines, and these are brought into relationship with some contemporary formulations of psychological problems. It seems a pity that the author hardly hints at the issues raised by 'noise', nor does he provoke the concept of probability to retaliation. Discretion, not valour is his watchword. The book is an excellent two shillings' worth which students of psychology should find will cheer but not inebriate.

R. C. OLDFIELD

L'Esthétique. By DENIS HUISMAN. Paris: Presses Universitaires de France. 1954. Pp. 128. Price not given.

This is a very interesting and illuminating little book, and should have a wide sale and be translated into English. Its author is secretary of the *Revue d'Esthétique* and of the *Société française d'Esthétique*. The book is divided into three parts. In the first part the author deals with the great steps or phases in the history of aesthetics: first, the Platonic or dogmatic age, as he calls it; secondly, the critical or Kantian stage, including pre-Kantian and post-Kantian aesthetics; thirdly the positivistic or modern age. In the second part of the book the author writes about the great fields of aesthetic inquiry: first the philosophy of art; secondly, the psychology; and thirdly, sociology of art. In the third part of the book he deals with special problems and aspects of art and aesthetics: first, the problems of the relationships of art with science, morality, nature, industry and religion; secondly, the classification and grouping of the various arts; thirdly, the aims and methods of aesthetics and its outlook for the future. In all it is a very comprehensive and widely based work, and the author concludes by saying that if aesthetics refuses to be experimental, precise and positive, it will cease to exist.

R. W. PICKFORD

Catholics in Psychology. By HENRYK MISIAK and VIRGINIA STAUDT. MacGraw-Hill Publications Inc. 1954. Pp. xv + 309. 40*s*.

This is a book of 300 pages and is one in the series of McGraw-Hill Studies in Psychology. The frontispiece shows a fine portrait of the late Cardinal Mercier, a pioneer in bringing Catholic studies into line with modern psychology, at the University of Louvain. Here was founded a brilliant school of experimental psychology where Michotte became Professor.

The authors start with a good résumé of the rise of psychology in general, pointing out that in Germany there were two main streams, one deriving from Wundt, and the other from Brentano—the founder of 'phenomenology'. The fact that the latter, working in the materialist climate of the nineteenth century, could not reconcile his faith with the new 'science' and gave up being a Catholic, may have been responsible in part for the distrust felt by some Catholic philosophers and theologians for this 'psychology without a soul' which was being torn from the sheltering arms of philosophy.

However, this rift did not long continue, and the cause of free psychology was ably served by Catholic psychologists in various parts of the world, notably Verner Moore—a Benedictine monk and also a psychiatrist, in America—and Agostino Gemelli, Franciscan friar, and also Rector of the University of Milan. Among many brief biographies, there is one on the late Sister Marie Hilda, that remarkable little nun of Glasgow, who started one of the first Child Guidance Clinics in this country, and directed it till her recent death.

'There is of course no Catholic psychology; there is only psychology', as the authors remark, and there is no necessary conflict whatever between the Catholic religion and science.

One wonders whether it was necessary to undertake this lengthy compilation—which is certainly very complete and accurate—in order to demonstrate that there are Catholics working in the psychological field, and that there is now much more understanding on both sides than there was a generation or two ago.

However, anything which breaks down misunderstanding and prejudice is to be welcomed, and the authors have ably carried out their task.

C. L. C. BURNS

Police Work with Juveniles. By JOHN P. KENNY and DAN G. PURSUIT. Springfield, Illinois: Charles C. Thompson; Oxford: Blackwell Scientific Publications Ltd. 1954. Pp. xxi+371. 55*s.* 6*d.*

This book was written as a practical manual for police men and women in the U.S.A. As might be expected part of it is concerned with elementary advice of a factual nature such as how to keep a file and the duties of a policewoman. However, there is much more to the book than this. It does tell the police in a clear and simple way some of the underlying causes of delinquency. Sound advice is also given on how and how not to deal with people, and how and how not to interview. Although the book contains nothing new for psychologists it is an interesting account of a very thorough and praiseworthy attempt to bring the behaviour of the police more in line with the ideas and precepts laid down by psychologists and other experts on delinquency.

MADELINE KERR

Three Men. By JEAN EVANS, with an Introduction by GORDON W. ALLPORT. London: Gollancz. 1954 (first copyright in 1950). Pp. xviii+297. 15*s.*

The author has given three vivid word pictures of delinquent, neurotic or problem boys and of their development. They are Johnny Rocco, probably a psychopathic delinquent, William Miller, probably an hysterical delinquent, and Martin Beardson, a partial homosexual. She describes their personal histories with skilful artistry based on a considerable understanding of clinical psychology, and attempts in a somewhat 'racy' literary manner to make them come to life in her descriptions. In this attempt she is very successful. She undertakes a kind of clinical biography in case-history writing, which would be fascinating and useful to the student of social sciences, of child guidance or the problems of adolescence. Interesting and illuminating though these biographies may be, they are very long and much of the detail in unnecessary. It would have been more useful to have had ten biographies suitably presented in an equally vivid and sympathetic though briefer form. Though the ordinary reader and student will find the novel-like writing very absorbing, he will probably be quite incapable of providing for himself the theoretical interpretation and synthesis which is very necessary and on which the biographies obviously depend though it is never presented. Nevertheless, the book would form an admirable basis for instruction in clinical psychology if the teacher could provide the necessary explanations.

R. W. PICKFORD

The Durham Research Review, No. 5, September 1954. London: University of London Press. Pp. 62. 5*s*.

The contents of this issue are divided between education and psychology. Three articles deal interestingly with byways in English educational history. J. Conway Davies describes, with frequent quotation from his sources, some Elizabethan schemes of education, and J. L. Dobson gives a parallel detailed picture of a school run on Dr Bell's system in the early nineteenth century. The feature of both schemes is the exacting demands on the pupil, and the apparent absence of any attempt to fit the curriculum to the pupil's interests and abilities. An article by M. Gilbert discusses the influence of the German reformers of the 1870's on language teaching in England, illustrating incidentally the transition from the educational attitudes of Dr Bell's time to that of the present day. All these articles are adequately documented.

Of more interest to the readers of this *Journal* will be the psychological contributions. Dr Macfarlane Smith describes the construction, analyses and validation of a Space Test. He confirms the evidence for a 'Space Factor' at 11 +, and finds the space test correlates more highly with school marks in Technical Subjects than does the conventional verbal group test. But the correlations are in general low, and the numbers not large. Dr Smith also speculates on the relationship of space-test scores to masculine and feminine traits. T. P. Howard, in the field of experimental psychology, examines the conditions under which the size-weight illusion operates, finding a change from positive to negative error as the weights used become smaller. The discussion raises a number of theoretical points which could profitably be pursued farther. Graham and Sluckin report on a pilot study of factors which workers consider important in their jobs. The inquiry is slight, but does not claim to be more.

The difficulty of editors of such periodicals as this is to maintain a supply of articles of adequate quality in competition with the more specialized journals. The heavy weights are normally attracted to the latter type of journal; but Prof. Smith and his co-editors can be congratulated on a set of competent and interesting light-weight contributions.

J. MAXWELL

The Teaching of Arabic Numerals. By G. G. NEILL WRIGHT. The Scottish Council for Research in Education. XXXVI. The University of London Press. 1954. Pp. 45. 2*s*. 6*d*.

This is a summary of *The Writing of Arabic Numerals* by the same author published in 1952 and reviewed in this *Journal* in February 1954. It is published by the Scottish Council for Research in Education in pursuance of their policy of making the results of investigations easily available to those who are more interested in the results themselves than in the steps by which they have been reached. Readers who wish to review the evidence on which the results rest are referred to the larger volume.

This very much abridged version is a very successful summary. It brings out clearly the importance of teaching the child a form or prototype for each numeral which he can write legibly, quickly, with ease, and as nearly as possible independent of attention. Advice on the selection of prototypes is given and many illustrations are provided.

For any one concerned with teaching children arithmetic this is a most useful little book.

JOHN SUTHERLAND

OTHER PUBLICATIONS RECEIVED

A. M. Blackwell. *Lists of Researches in Education and Educational Psychology: Supplement I.* (Publications of the National Foundation for Educational Research in England and Wales, No. 7.) London: Newnes Educational Publishing Co. Ltd. 1954. Pp. 57. 10*s.* 6*d.*

Albert Burloud. *Psychologie de la Sensibilité.* Paris: Libraire Armand Colin. 1954. Pp. 224. 250 francs.

Egon Freiherr von Eickstedt. *Atom und Psyche.* Stuttgart: Ferdinand Enke. 1954. Pp. 158. Paper, DM. 12.00; linen, DM. 14.20.

Peter Fireman. *Perceptualistic Theory of Knowledge.* New York: Philosophical Library. 1954. Pp. xi + 50. $2.75.

A. Michotte. *La Perception de la Causalité.* Anvers (Belgium): S. A. Standaard-Boekhandel; Second edition, 1954. Pp. viii + 306. Paper, 175 Belgian francs; bound, 220 Belgian francs. (A reprint of the original text, with a new 2-page preface and a short list of relevant research published since the appearance of the first edition (1946).)

John Penman and Marion C. Smith. *Intra-Cranial Gliomata.* (Med. Res. Council, Special Reports series, no. 284.) London: H.M.S.O. Pp. vi + 70. 1954. 5*s.*

Calvin P. Stone. *Abnormal Psychology Glossary* (revised edition). Stanford University Press; London: Geoffrey Cumberlege. 1954. Pp. 24. 4*s.* net.

British Film Institute. *Films on the Sciences.* London: British Film Institute, July 1954. 1*s.*

J. Campbell (ed.). *Spirit and Nature: Papers from the Eranos Yearbooks.* London: Routledge and Kegan Paul. 1955. Pp. xvi + 492. 36*s.*

Mirclea Eliade. *The Myth of the Eternal Return.* (Translated from the French by W. R. Trask). London: Routledge and Kegan Paul. 1955. Pp. xi + 195. 18*s.*

J. J. Poortman. *Ochêma (History and Import of Hylic Pluralism).* Assen, Holland: Van Gorcum. Pp. 179 (incl. 8-page Synopsis in English). Paper Fl. 6.90, cloth Fl. 8.25.

W. Bromberg. *Man above Humanity: a History of Psychotherapy.* London: Pitman Periodical Publishing Co. 1954. Pp. xiii + 342. 45*s.* net.

D. McLean. *Nature's Second Sun: Leaves from a Teacher's Log.* London: Heinemann. 1954. Pp. xii + 130. 10*s.* 6*d.* net.

J. C. Flugel. *Man, Morals and Society.* Harmondsworth: Penguin Books Ltd. 1955. Pp. 415. 3*s.* 6*d.* (With a new Preface.)

J. B. Rhine. *The Reach of the Mind.* Harmondsworth: Penguin Books Ltd. 1954. Pp. 198. 2*s.* 6*d.*

G. Cuisenaire and C. Gattegno. *Numbers in Colour: A new Method of Teaching Arithmetic in Primary Schools.* London: Heinemann. 1954. Pp. viii + 39. 4*s.* 6*d.*

A. A. Congalton (ed.). *Hawera—a Social Survey: A report of a Community Venture.* Hawera, New Zealand. 1954. Pp. xv + 219. 10*s.*

A. J. Nixon. *Divorce in New Zealand: a Preliminary Survey.* Auckland University College (Bulletin No. 46, Sociology Series no. 1). 1954. 4*s.*

Charles Eisenmann. *The University Teaching of Social Sciences: Law.* UNESCO. 1954. Pp. 133. 6*s.* net.

EDITORIAL NOTE

Intending contributors are asked to note that changes have been made in the instructions on the *Preparation of Papers for the British Journal of Psychology* which appear on the inside front cover.

The principal alteration is that it is recommended that in future each article should be preceded by a short abstract or summary, conveying the gist of the article and embodying the principal conclusions. This should, to some extent, render the inclusion of a formal summary on a similar scale at the end of an article redundant, although a concluding section or paragraph summarizing the argument or drawing attention to the main findings will not be out of place. The guiding principle should be to regard the summary as being quite separate from the article proper, which in its turn should not depend in any way for comprehensibility or completeness on the summary.

The summary, placed at the beginning of the article, will also replace the list of numbered headings and sub-headings which at present appears in this position. It is felt that such a list is unnecessary for articles of a length falling within the present recommended maximum. Articles should, however, still be subdivided in the way described.

It is hoped in future to publish a small proportion of very short articles (under 1000 words), such as Apparatus Notes, Research Reports, etc. It may occasionally be possible, as space allows, to accord priority to such contributions. Short papers of this kind may not require to be subdivided; a summary may also be unnecessary.

All book reviews will in future be signed. Reviewers are reminded that full bibliographical data should be given, in the following order: Title; author; place of publication; publisher's name; date; number of pages (with preliminary pages noted separately); price. Where a publisher acts as agent for a foreign house this should be mentioned.

The use of roman numerals in the citation of volume number references is being discontinued.

The new style will come into effect with Volume 47. Since it will not be possible to include in the present volume articles other than those already received, all contributions submitted from now on should conform to the altered requirements.

ERRATUM

The authors of the article "The effect of hunger upon drinking patterns in the rat" (A. D. Calvin and R. A. Behan, this *Journal*, XLV, Part 4) have intimated that the following footnote to their title was inadvertently omitted:

"The authors are indebted to E. F. MacCaslin whose observations prompted this experiment."

VOLUME XLVI AUGUST 1955 PART 3

ON THE LIMITS OF EXPERIMENTAL PSYCHOLOGY[1]

BY B. A. FARRELL

Wilde Reader in Mental Philosophy, University of Oxford

Wittgenstein has said: 'The existence of experimental methods [in psychology] makes us think we have the means of solving the problems which trouble us; though problem and methods pass one another by askew [windschief aneinander vorbei laufen].'[2] As far as I can discover, Wittgenstein does not produce anything to support this contentious and obscure remark. I shall use it as my text, and try to indicate where it is useful and where it is quite unhelpful. In doing so I shall take you for a logical walk (or, more probably, a trot) and resume the topic of the interesting but, to me, inconclusive symposium on Experimental Psychology which was held at the Annual General Meeting at Liverpool a few years ago.

TWO FEATURES OF PSYCHOLOGICAL WORK

I begin by reminding you of two features that characterize the results of psychological endeavour.

In the course of the past eighty years or so, psychologists have discovered a great deal about what might loosely be called 'the adaptive behaviour' of organisms. These discoveries are to be found in the thick forests of the literature. What psychologists have not yet discovered is how to synthesize or unify their discoveries. That is to say, they have not succeeded in hitting on some generalizations by means of which, to put it loosely, they will be able to explain the mass of varied and particular discoveries they have made. For the generalizations that express the discoveries they have made are relatively low level, or specific, in character. What they have not yet succeeded in doing is to find sufficiently high-level generalizations, by means of which they can derive, roughly speaking, their specific generalizations and the particular statements embodying their discoveries.

Well, this is a familiar, indeed a commonplace, feature of psychological endeavour. The second and related feature I wish to mention is perhaps not so clearly appreciated, but, since it takes some time to describe it adequately, I shall merely indicate what it is. As we all know, it is characteristic of many psychological generalizations that we can find exceptions to them. It is notorious how vulnerable they are apt to be. The second feature, therefore, I wish to mention is the vulnerability of generalizations in psychology. It is obvious that these two features of psychological work are closely related, but I shall not pause to examine their relationship here.

EXPERIMENTATION AS THE SOLUTION

How, now, can this present situation in psychology be remedied? Psychologists have taken it for granted that this situation *will* be remedied in the long run. In other words, they have presupposed that remedial action is possible *in principle* and that, consequently, there is nothing absurd in supposing that we will one day light upon the high-level generalizations which will unify our psychological knowledge, and replace generalizations which are vulnerable by others that are in effect definitive. This presupposition is not obviously true, and I am not even sure that I know what it means. But I shall accept it for the purposes of this paper and pass on.

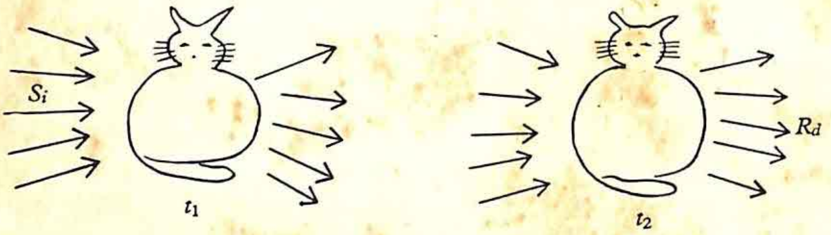
---

[1] This article is based closely on a paper read at the Annual General Meeting of the British Psychological Society, at Nottingham, April 4th, 1954.

[2] Wittgenstein (1953), part II, xiv. This unsatisfactory translation is my own.

So the question is: How can this situation be remedied in *practice*? What psychological method, or family of methods, will be sufficient to do so?

When one stands on the high shores of this island and breathes in the transatlantic winds that beat upon them, one gets the impression that some psychologists take the line that what is sufficient to remedy the situation are the standard, or typical, procedures of psychological experimentation, when executed in the usual context of what is commonly called the hypothetico-deductive method. By the typical procedures of psychological experiment, they mean, I think, to refer to something quite orthodox. They mean to refer to a type of experimental design that can be represented in the following sort of text-book diagram (after Andrews, 1948, chap. I):



Here we consider an organism during any one sufficiently short slice of its life history, t_1. The organism is thought of as being bombarded, or disturbed, by a large number of stimuli, or stimulus conditions, or what you will. It is also thought of as making various adjustments or responses at this time. Now some special condition of disturbance is introduced at t_1, those consequences we are interested in. This is labelled the independent variable, S_i. We then set about discovering what response, R_d, or whether some response R_d, at some later phase of the organism's life history (say t_2 or t_4) is a function of this independent variable or not; or in what way it is a function of it. In the course of doing this we are weighed down with care trying to keep everything else constant, arranging for controls, and so forth. Our overall concern in this sort of experiment is with connexions between disturbing (or stimulus) conditions on the one hand, and behaviour, or conduct, or performance on the other.

Of course, this picture gives a misleading impression in various ways of the typical or standard procedure of psychological experiment that many psychologists have in mind. For example, it suggests that the stimulus conditions always come from the external environment and that the dependent variables are necessarily overt or publicly observable performances. Still, I hope the picture is good enough for my purposes. What is more important is to notice what the picture excludes. It excludes the use of differential methods. It excludes much of the work done in what is loosely called 'Physiological Psychology'. Thus, much of the work reported in chapters II to VIII in Stevens's *Handbook* (Stevens 1951) would not pass. For work on 'Excitation and conduction in the Neuron', 'Synaptic Mechanisms', 'Sensory Mechanisms' and 'Motor Systems'—such work does not consist typically, if at all, of discovering how some organismic performance is connected with some certain independent variable. Clearly, too, this picture of the standard psychological experiment excludes the typically clinical investigation, designed to discover, for example, how lesions of the parietal lobes are connected with disturbances of spatial perception, and so forth.

I have said one gets the impression that some psychologists seem to take the line that the typical procedures of psychological experimentation, in the context of the hypothetico-deductive method, will be sufficient to remedy the situation. I won't bore you with supporting quotations (but see Hull, 1943; Skinner, 1953); instead I will go on at once to ask: Is this belief well founded? This question, and all questions of this sort are notoriously difficult to answer for the well-known reason that we cannot tell in advance what scientific strategy will do the trick. If one advocates a certain type of scientific strategy, one is in effect backing a certain horse to win. What, therefore, we are doing in answering this question is to say whether it is reasonable, or sensible, to back the horse of experimental psychology or not. I shall set about answering this general question in the following way. First, I shall ask and try to answer this query. What do high-level generalizations look like where they have been achieved, and where, therefore, our knowledge in some particular field has been unified in some large or considerable measure? Is there any general feature that these high-level statements characteristically exhibit? I shall then ask a second question. Are there any reasons to suppose that the high-level generalizations we aim at in psychology will *not* exhibit this feature also? If there are no such reasons, if, in other words, psychology seems to be analogous to other sciences in this respect, I shall go on to a third question: Is experimental method in psychology, as typically depicted and used, likely to give us the high-level generalizations we require? If it will, then the belief of our transatlantic colleagues will have good support. If it will not, then we will have a good reason for rejecting their belief.

HIGH-LEVEL GENERALIZATIONS IN SCIENCE

I begin with the first query. What do high-level generalizations look like where they have been achieved? Here I shall speak very ignorantly and tentatively, and I shall have to rely a good deal on what others tell me.

The first thing to emphasize, I think, is that the high-level generalizations which have been achieved do not all look alike. Consider, for instance, those which were achieved in Mechanics. These are generalizations about physical things like levers, pulleys, freely falling bodies, the moon and the planets and the like; and they state certain invariances that these macroscopic objects exhibit. This sort of generalization is important for us, all other reasons apart, because some psychologists appear to have been influenced by the example of progress in Mechanics, and to have supposed that the high-level generalizations of psychology will also be of this sort (e.g. Hull, 1943, chap. II). Now, *if* they are going to turn out to be of this sort, then a good case can be made out for saying that experimental method will be sufficient to give us them. But I think a good case can be made for the view that the high-level generalizations of psychology will *not* look like this at all, and that the example of Mechanics is quite misleading. There are three obvious reasons for holding that this example is misleading: (i) There are the wide differences between the phenomena exhibited by the sun and the moon, etc., and the phenomena of adaptive behaviour.

(ii) The generalizations of Mechanics are, perhaps, uncharacteristic of those usually achieved when a science progresses. (iii) The high-level generalizations which have been achieved in fields more closely related than Mechanics to the field of adaptive behaviour are of a different sort to those found in Mechanics. Consequently, when I now go on to say what the high-level generalizations of science look like, I shall pick on and stress a type different from that found in Mechanics. I pick on this different type because it does seem to be more *characteristic* of the generalizations achieved in the course of scientific progress, and because it is the sort of generalization that has been achieved in fields not too far removed from that of adaptive behaviour. I shall introduce this different type in the following way.

When one looks at the history of science, it seems a rather striking fact that progress is often rapidly achieved when the workers in any given field are able to hit upon generalizations which state invariances between minute things (or constituents) or classes of things (or constituents), or state an invariant property of some minute things or classes of things (or constituents). Thus, chemistry is transformed as soon as it hits on the minute things and their properties upon which the phenomena of combustion and chemical combination depend. Biology in general gets under way when it can produce generalizations about microscopic entities, called 'cells', and explain the phenomena of reproduction, growth, metabolism and the like by means of the properties of these, and lesser, microscopic entities. Physics has managed to unify large parts of its field in the same way; and, where it has not yet done so, we are strongly tempted to say that this is because it has not yet hit upon the fundamental particles and/or their properties necessary to achieve the synthesis required. In short, it is a striking and fairly common, but by no means universal, feature of the high-level generalizations in science that they are concerned with the minute constituents of macroscopic objects or things, and with the properties of these minute constituents.

This is all a familiar, indeed a commonplace story. But it is also a vague story as it stands, and it is necessary to make some of it a little more precise. For this purpose I shall use some distinctions related to some suggested by Broad (1933, chap. XIV, § 3).

Consider what would correctly be called 'a thing' or 'an object', or 'a sample of a particular thing or substance'. This thing can be treated as a *simple* or as a *compound* thing. Thus, when I occasionally make the morning porridge, I treat the salt I put into the saucepan as a *simple* thing, in contrast with the porridge which it goes to make up. But when a chemist considers a given quantity of salt, he regards it as a *compound* thing, consisting of two other sorts of things or constituents, namely sodium and chlorine ions related in a cubically structured lattice. To this chemist it is the atoms of sodium and chlorine, with the respective positive and negative charges, that are the *simple* constituents of salt.

Next, a compound thing may have dispositional properties which the simple things composing it do not have. For example, common salt is salty to taste, but sodium and chlorine are not. A piece of sodium at room temperature is solid, but a sodium atom is not. Following Broad, I shall call a property which belongs to a compound thing as a whole, and not to any of its constituents, a *collective* property. A collective property may be of one of two sorts, which I shall explain by a crude and inexact example. Suppose I find some pieces of Meccano lying round the playroom. Suppose I know the dispositional properties of these different pieces, for example, that they are rigid, that they are attachable to each other in various ways, etc. Suppose, next, that I consider putting the pieces together in a certain special way. They will then form a compound thing, with possibly some collective properties, for example, that the compound thing is able to move smoothly along a level surface carrying small loads—in other words, is able to act like a lorry. Now, it may be possible for me to infer, or deduce, this collective property from my knowledge of the separate pieces of Meccano, and from my knowledge of the special relations which I am supposing these pieces have to one another in order to form this particular compound thing. In other words, consider the generalization 'All compound things, of which this is an instance, have the collective property, viz. of being able to move smoothly, etc.'. This generalization follows logically from the generalizations I have at my disposal about the separate pieces of Meccano and about the properties of these pieces in the special mutual relations considered. Where this is the case, I shall call the collective property, following Broad, 'a reducible property', These contrast with collective properties which cannot in principle be deduced from a knowledge of these constituents in the special relations involved in making up a particular compound thing. Broad calls these 'emergent properties'. For example, common salt is 'salty', but neither of its constituents are; and it can be argued that this property, 'salty', is not deducible in principle and therefore is emergent.

One more piece of logical tooling, though I am afraid it is a very rough piece. Consider and contrast a hydrogen atom in a molecule of water, and, say, a leg of a live human being. It is in fact the case that hydrogen atoms can, and do, normally exist in nature in special relations other than those they are in when in water. By contrast, human legs cannot, and do not, normally exist in nature in special relations other than those they are in as part of a human organism. They only have a separate existence in special and infrequent contexts, like those of the anatomy room, the forensic medicine laboratory, and the like. Now, when a hydrogen atom is not standing in the special relations involved in a water molecule, it would be normally natural and true to say of it that 'it is able to enter into these special relations', or that 'it is able to combine with another hydrogen atom and an oxygen atom in a certain way to form a water molecule'. By contrast, since a human leg does not normally occur in nature outside of its special relations in an organism, it is not normally natural and true to say of it that 'it is able to enter into these special relations'—'that it is able to combine with other parts of the body to form a human organism'. I shall call the sort of part, of which the hydrogen atom is an example, an 'independent' constituent of a compound, in this case water. The sodium of common salt and the pieces of Meccano toy are other examples of independent constituents of their compounds. I shall call 'a human leg' a 'dependent' constituent of the compound which is the human organism. Many of the constituents of biological compounds appear to be of this latter sort. Of course, this is a rough distinction, as I have already indicated. To speak of what can, and does, normally exist in nature outside of special relations is to speak vaguely; and what it is normally natural and true to say of a constituent of a compound will differ under different circumstances. However, this is all I have the time for, and I think it is good enough for my purposes.

The relevance of the distinction is this. I defined a reducible collective property as one that was deducible from, in part, statements about the properties of constituents when they are *not* in the special relations in which they do stand in the compound thing. From this it follows that, when the constituents concerned are not in these special relations, they must still be able to function with the other constituents of the compound thing to exhibit the collective property we are deducing. But the only constituents which retain this property when not functioning as part of a compound are the 'independent' constituents of this compound. That is to say, I defined the notion of 'reducible property' by means of the notion of the 'independent constituents' of the compound in question. If, then, one tries to reduce the collective properties of biological compounds, one has to be careful to find independent constituents for the purpose; and this may prove much more difficult and tiresome than has been the case with the collective properties of physical things.

In so far, now, as we have unified our knowledge of nature, we have apparently done so by the use of a few, different sorts of fundamental particles, or minute things, or constituents, and a few properties of these minute constituents. The high-level generalizations we use for this purpose express invariances about these constituents and their properties. By means of them we derive a large part of the remaining generalizations of physics and chemistry, and we thereby explain some of the wide variety of physical phenomena we confront. So when the schoolboy asks us: 'Why is salt soluble in water?', we answer him by deriving this collective property (of solubility in water) from the properties of the constituent ions of salt and molecules of water. We answer him by trying to show how the solubility of salt in water is a reducible property. When he asks: 'But why does it reach a saturation point? And why has the temperature of the water got anything to do with it?', we try to explain the phenomenal variety that salt-in-water can show by reference to the properties of the minute constituents concerned under different conditions.

HIGH-LEVEL GENERALIZATIONS IN PSYCHOLOGY

I turn now with relief from logic and general considerations to the problem about psychology. I have tried to answer the question: 'What do high-level generalizations characteristically look like where they have been achieved?' The next question we set ourselves was this: 'Are there any reasons to suppose that the high-level generalizations we aim at in psychology will be any different in character?'

If they are to be of the *same* sort as those already and characteristically obtained elsewhere, what will they look like? What special form, or forms, will they take? It is clear at once that, if psychologists are to carry out the same sort of programme that has been successfully executed in other fields, they will have to regard 'an organism' in the same sort of way in which it is generally regarded in the biological sciences. They will have to regard it as a compound thing composed of a vast mass of minute things called 'cells', which in turn have other minute things in relations composing each of them. Now the psychologist is specifically concerned with discovering generalizations about the adaptive behaviour, or performances, of organisms. In doing this he is concerned with the various capacities or powers of organisms—the capacity to learn, to remember, to imagine, to see, to solve problems, to tolerate frustration to some extent, to exhibit goal-directed activity, and so on and so forth. It is usual to call these powers or capacities 'dispositions' or 'dispositional properties'. The examples I have listed here are, in an important sense, more or less generic dispositions of all organisms or of various species of organisms with which psychologists are concerned. They contrast with the more specific disposition a rat acquires when it has learned to solve a particular maze, or with the one I acquire when I learn the etymology of the word 'Psychology'. In the course of studying 'adaptive behaviour', psychologists have to handle both the generic and the specific dispositional properties that organisms can and do exhibit. The next point is an obvious one. It is clear that these properties belong to an organism as a compound thing, and not to any of its constituents. Thus, it is not the brain which has the capacity to learn; it is the organism which can do this. It is not any 'cell assembly' which has the specific dispositional property of being able to solve a particular maze; it is an organism that has this property. It is not the occipital lobe which has the power of sight; it is an organism. So in being concerned with certain dispositional properties of organisms, psychologists are concerned with certain *collective* properties which they exhibit.

If, then, psychologists are to unify our knowledge of organismic performance in the same sort of way that has been characteristically achieved elsewhere, they will have to show that the collective properties of organisms with which they are concerned are *reducible* properties. For this purpose they will have to find a few sorts of constituents of organisms and a few fundamental properties they possess—constituents and properties sufficient to enable them to derive the whole range of collective properties that organisms exhibit. Our knowledge about these few constituents and their properties would be contained in a few high-level generalizations expressing the known invariances about these constituents and their properties. From these high-level generalizations we should have to be able, roughly speaking, to deduce, and thereby explain, the wide variety of organismic performance we do confront.

What would a reduction of these collective properties look like? It is only possible at present to give the general form of such a reduction. For we are held up at the outset by not even knowing what candidate, or candidates, to accept for the role of the minute constituents we require. The popular candidate, I suppose, is 'the nerve cell'. Let's adopt it for the present (I shall raise an obvious objection in a moment). Consider the collective property of, say, intelligence, and let me introduce you to two charming white rats, Bertie and Sammie. Suppose we had discovered the properties of some or all nerve cells from which, after investigating the nerve cells of Bertie and Sammie, we could deduce that Bertie was more intelligent than Sammie. We would then be said to have reduced the collective property of 'intelligence'. Or consider some generalization, or tendency statement, about 'intelligence'. For example, 'Rats brought up in good homes, with books, toys and plenty of exposure to intelligent adults tend to be more intelligent than rats brought up in bad homes' (cf. Hebb, 1949, chap. xi). If we knew what the properties of the nerve cells were that determined this difference between the two classes of rats, we should then have available high-level generalizations, which would enable us to explain this fact about differential experience tending to produce differential adult intelligence in rats. Moreover, this difference is more marked with some rats than with others, and we should require our generalizations to explain this fact by reference to differences between the properties of the nerve cells of Bertie, of Sammie and of their relatives.

I can now answer the second question we set ourselves: Are there any reasons to suppose that the high-level generalizations we aim at in psychology will turn out to be of this sort or not? Let us bear in mind that I only ask this question of 'individual' not of 'social' psychology, where quite different sorts of considerations may arise.

There would be good reason for saying: 'No, the high-level generalizations we require will *not* be of this sort', if we had independent evidence for supposing that all the plausible candidates for the role of minute constituents had characteristics that excluded them from playing this role. For example, if we knew that the nervous system merely acted as an automatic and perfectly regular purveyor of impulses from stimuli to responses, and that nerve cells exhibited no autonomous activity of their own and could not do the job of storage—if we knew that this sort of thing was true of the nerve cells, then these characteristics would probably preclude them from being the minute constituents we want. But, as I understand the position, and my understanding is limited, we have no such knowledge about the nerve cell or any other plausible candidates, if there be any. So I conclude that we have no good reason to suppose that the high-level generalizations we require will *not* take the form I have described. Indeed, it is noteworthy that psychologists who are interested in *how* organisms manage to do the things they do, quite naturally start looking, directly or indirectly, for the parts, or constituents, of their bodies which enable them to perform in this way.

But I think we have *some* reason to suppose that the high-level generalizations we require will not employ 'nerve cells' as the, or one of the, fundamental constituents. For I have already pointed out that the minute constituents or parts we need for reduction have to be independent constituents, and not dependent ones. But 'a nerve cell' appears to be a dependent part, in the circumstances likely to concern us. Consequently, we will not succeed in reducing the collective properties of organisms by means of nerve cells. This suggests that we will not obtain the high-level generalizations we require if we use them. While concentration on the nerve cell will give us something, possibly a great deal, it looks as if all that it *can* give us will be insufficient. It is too gross and dependent an object. However minute a theorist like Hebb is, his fault may be that he is not minute enough (Hebb, 1949). It looks as if we will have to plunge into biochemistry and electrophysiology. I find this a sobering thought. It also looks as if the work of constructing machines which duplicate organismic activity may have a special use in drawing our attention to the *sort* of minute parts of organisms which will be able to play the role of the independent and fundamental constituents we want.

THE INSUFFICIENCY OF PSYCHOLOGICAL EXPERIMENTATION

I shall now apply these remarks to answer the third question I asked: Is experimental method, as typically depicted and used, likely to give us the high-level generalizations we require? If we only use this typical or standard procedure, we seem to be restricted, in principle, and briefly speaking, to two types of generalization. (A) Those connecting stimulus conditions with responses. (B) Those connecting *either* stimulus conditions with some hypothetical internal states of, or occurrences in, the organism, *or* some hypothetical internal states or occurrences with responses, *or* some of these hypothetical states or occurrences with others. I think we have some reason for holding that neither of these two sorts of generalization, A or B, are likely to do the trick.

Consider type A. The words 'stimulus' and 'response' (and their alternates) are admittedly slippery customers, and I am not at all clear how they are used in the literature. As a first move, however, I suggest that they are *not* used to stand for 'constituents' of organisms or other things. The word 'stimulus' is perhaps usually used to refer to some occurrence (or family of occurrences) in the history of other things; or to some feature or set of features exhibited by some thing. This use is good enough to cover the standard cases of buzzers, electric grids, Skinner boxes, tachistoscopic material, and similar amusements. The word 'response' is usually used perhaps to refer to some occurrence in the life history of an organism, in whatever way this is carved out of its relatively continuous performance. So, however 'stimulus' and 'response' are used in the literature and technical discourse, they do not seem to be used to stand for the minute constituents of organisms, or even for the properties of such constituents. Moreover, since stimuli and responses are indefinite in number and indefinitely variable in character, they will not give us the few fundamental constituents with a few properties which we want. S–R generalizations, therefore, do not appear to be the sort we require.

Consider the second type, B. These generalizations look more promising, since they include something other than stimuli and responses, namely, the internal states of, or occurrences in, the organism; and these states or occurrences may be instances of properties of the sort we want. But our hopes are dashed when we look at some of the internal states and occurrences which have been suggested. Take 'Habit strength'. We are told that this 'exists as an organization, as yet largely unknown, hidden within the complex structure of the nervous system' (Hull, 1943, chap. VIII). This, then, is a property of a large part of the body, namely the nervous system and is far too gross to do the job required. The same applies to internal states or events like that of 'afferent neural interaction'. Alleged occurrences like 'Movement-produced Stimuli' (Guthrie, 1935, *passim*) are rather different. They are manifestations of the properties of certain constituents, namely and in particular, the skeletal muscles, tendons and joints. However, we do not seem to have any evidence to suggest that these are sufficiently fundamental for reductive purposes.

It is worth adding, of course, that there is no reason *in principle* why an experimentalist should not hit upon generalizations which connect, say, stimulus conditions with constituents of the organism which *are* sufficiently fundamental for reduction. There is no reason in principle why he should not postulate the right sort of constituents and properties. The snag is that as long as he sticks to the techniques of experimental psychology, he is unlikely to be in a position to maintain empirical control over statements about these postulated constituents and their properties—a control which Hull rightly insisted he should maintain. He will be unable to test his generalizations, and they will either pass into the sceptical hands of neurologists, or else be relegated to the library of science fiction.

The answer, then, to our third and last question is in the negative, and follows quite naturally from what has gone before. The use of experimental methods alone, as typically depicted and used, will not be sufficient to remedy the present situation and give us the high-level generalizations we want. The confidence of some of our American colleagues in these methods seems to be overdone. At the beginning of this paper I mentioned two related problems that confronted psychologists—the absence of the high-level generalizations needed for unification and the vulnerable character of many of the generalizations they do employ. In so far as psychologists are concerned to deal with these questions, it is true to say that experimental methods, in the typical sense, miss the problems of the psychologist. Problems and methods here do pass one another by askew. In this important respect Wittgenstein seems to have been right. Further, I suspect Wittgenstein may be right in implying that the mere existence of experimental methods makes psychologists think that they have the means of solving the problems which trouble them, including those of unification and vulnerability. But at present I am not clear enough about this implication to pursue it any further.

THE NECESSITY FOR PSYCHOLOGICAL EXPERIMENTATION

Now let me turn briefly to the other side of the picture. Where and how is Wittgenstein's remark unhelpful? The first thing to point out is that the important respect in which Wittgenstein is right is also a very limited respect. Psychologists are in fact concerned with all sorts of *other* problems than the two I have been considering. I suspect that it would be a gross falsehood to say that most of them were actually concerned with this problem at all. It is only ignorant people, like philosophers, who think that psychologists are in active and frantic search for the balm of fundamental law. It is quite unnecessary to draw the attention of a psychological audience to the varied character of psychological problems. Because the typical methods of experimentation may not throw much light on the problems of unification and vulnerability, it does not follow in the least that they throw no light on the *other* problems of contemporary psychology. It does not follow that they pass *these* problems by. Moreover, it is not the case that experimental methods are all of *one* sort. Thus, when I distinguished the typical, or standard, form of such methods, I pointed out that much of the experimental work in chapters II–VIII of Stevens's *Handbook* would not qualify as typical. Wittgenstein's remark suggests that the problems confronting psychologists are all of one sort, and that the experimental methods they use are also all of one sort. This is a very misleading suggestion, and in these respects his remark is quite unhelpful.

But Wittgenstein's dictum is defective in another and more serious way. There is, I have argued, some reason for holding that experimental methods in psychology, in the context of the hypothetical-deductive method, are not *sufficient* to unify our psychological knowledge and give us the invulnerable generalizations we want. But I have not argued that experimental methods are unnecessary for this purpose. On the contrary, I think a good case can be made for saying that experimental methods, while not sufficient, are *necessary*, in some sense, to enable us to unify our psychological knowledge and achieve invulnerability. Wittgenstein's remark is defective, since it makes us overlook this possibility. To say that problems and methods in psychology pass each other by 'windschief' is to suggest that they miss each other *completely*—like two points travelling in different planes—and this is wholly misleading. As far as I can judge, psychologists themselves are apt to be uncertain and unclear about the way in which experimental methods are necessary at this point. Consequently, I shall now try to bring out shortly just why and how they are necessary.

When a chemist claims that the solubility of salt in water has been reduced, he produces a statement like this: 'Sodium and chlorine ions are separable in water.' He uses this statement along with others to explain why salt is soluble in water, and to predict that other substances will or will not be soluble also. I shall call this statement 'p', and the statement that 'Salt is soluble in water', statement 'q'. If now a psychologist claims to reduce the 'intelligence' of, say, rats, he would have to produce a statement of the following sort, say: 'Bertie's cortex is capable of carrying a neural load of sort ϕ to the amount F, and this temporal maze imposes a neural load of this sort ϕ to a lesser amount than F.' Let's call this statement 'p''. He will have to use p' to explain why, or predict that, 'Bertie is intelligent enough to solve this temporal maze'. Let's call this latter statement 'q''.

What is the relation between p and q, and between p' and q'? It would seem that in ordinary discourse p and q are not *logically* equivalent to each other, and that p' and q' are not either. Obviously, because when we talk correct English and say that salt is soluble in water, the rules of the use of the word 'soluble' contain no reference to ions, molecules and the like. Similarly, when we say that Bertie is intelligent enough to solve this temporal maze, the rules of use of the word 'intelligent' contain no reference to neural loads of sort ϕ, or to the cortex or to the brain. Well, part of the relation which holds between p and q, and between p' and q' seems to be this. If p is true, then q is also true; and if q is true then p is also true. I shall call this 'equivalence' if anyone wants a name for it, and write it '$p \equiv q$'. The same applies to p' and q'. Thus, our neurological reductionist, in search of the cerebral basis of intelligence (as this is usually put), requires his p' and q' to function in such a way that: if it is true that 'Bertie's cortex is capable of carrying a load ϕ to amount F etc.,' then it is in fact also true that 'Bertie has the intelligence to solve this temporal maze'; and conversely.

Now for the point of all this. It seems clear that for a psychologist to *reach* p', it is not *logically* necessary that he should observe, experimentally or otherwise, the behaviour of Bertie, or any other rat, in a temporal or any other maze. As far as I can tell, the ionic theory of 'solubility' was not reached by making further observations on salt and other substances dissolving in water. It was reached—as far as I can gather—by inference from the theory of crystal structure and the electro-magnetic theory of matter. Similarly, it is quite conceivable that a psychologist could infer from a certain general theory about the cortex and the load of information imposed by environmental situations of different types that rats will be able to solve this temporal maze. But in order to reduce 'intelligence', he has to *show* that $p' \equiv q'$. In other words, to put it roughly and briefly, he has to verify this general hypothesis that if any rat has such and such a cortical structure, then it is able to solve a temporal maze, or, as we would ordinarily and correctly say, that it has the intelligence to solve it. Conversely, he has to show that if a rat has the intelligence to solve this maze, then its cortical structure is of such and such a sort. But it is difficult to conceive of him showing all this without using the typical procedures of psychological observation and experiment. So it seems to be logically necessary that he use them for the purposes of verification.

But more than this. Before a psychologist or neurologist, or any one else, can *reasonably* claim that p' represents a reduction of 'intelligence', he has to have evidence that the manifestations of intelligence, which he claims he can explain by reduction, do correspond roughly with what we normally regard as being manifestations of intelligence. But he can only have this evidence if he, or others, have already carefully explored the manifestations of the collective property of 'intelligence'. How can we reasonably claim to be able to reduce the intelligence of rats (and so find its cerebral basis) if we do not know what rats are capable of doing? And how can we discover this at present except by psychological observation and experiment on rats? Such work is a necessary exploratory step in a programme of reduction. (Just the same, incidentally, is true of 'solubility'. Here the necessary exploratory work on the behaviour of salt in water was done before the theory of crystal structure, etc., was developed; and we are apt to overlook that the same stage was necessary with 'solubility' as is necessary at present with the collective properties of organisms.) Further, though it is possible in principle for a neurologist to light upon a p' without looking at the performances of rats or humans or any other organisms, I think it is a little unlikely that it will be reached like this. At present we are all very ignorant about the general functioning of the central nervous system. The work of psychologists, experimental and otherwise, is full of suggestions about the way in which the central nervous system must function if organisms are to do the things they can and do do. I think that reductionists will need to make grateful use of all these suggestions, and thereby help to have their attention drawn to the fundamental constituents and properties they require. So experimental work in psychology will also be *practically* necessary in order for us to reach our p' statements, and thereby reduce the collective properties of intelligence, vision, learning and so forth.

I hope I have done something to bring out why and how experimental methods are necessary in our search for unity and invulnerability. And I hope I have done enough to show just how defective Wittgenstein's dictum really is.

Concluding suggestions

Some concluding points. First, I want to draw attention to some implications of my remarks. *If* psychology achieves its high-level generalizations in the way so characteristic in the history of science, then it seems to follow that much theorizing in psychology over the past thirty years has been misplaced.[1] Much of it has assumed that experimental psychology is strong enough to arrive at Laws of Behaviour—laws which will sit on a base of, at present, unknown neurological generalizations, but which will be synthesized with the latter in due course as these become known. If the analogy holds between the development of psychology and that of the other sciences, then this assumption may be a mistaken one. There may be no laws of behaviour sitting upon a neurological base. The high-level and invulnerable generalizations of psychology may turn out to be quite different in character. It follows, further, that, however great or small the defects in the systems of a man like Hull, no improvements will ever enable them to do the big, grandiose trick we would like them to do.

What sorts of psychological work will probably make the most relevant contributions to the particular problem of high-level generalizations? I tentatively suggest the following: (1) Work in what is loosely called physiological psychology. This extends from work of the Eccles-Adrian sort to that of the Lashley type, and beyond. (2) Clinical work of both the diagnostic and surgical types. If, for example, one investigates the deteriorated performance of people with cortical lesions, one can make suggestions about the way in which parts of cortex and the brain contribute to the normal performance. When one does this, it is not only the case, as Zangwill has rightly said (at Liverpool), that one studies performance 'with the internal environment simplified' in various ways. It is also the case that one is keeping one's feet on the hopeful ground of things and their properties. And it may be much more important to do this than to make a fetish of conforming to the criteria of rigorous experimentation. (3) There is the psychological work which is designed and conducted to throw light on the general plan, or plans, according to which the minute constituents of our bodies mutually co-operate to do their job. I am thinking of work like that by Deutsch (1954). For reasons I have already mentioned there is no need whatever for psychologists to be apologetic about this work, and to grovel in the presence of neuro-anatomists, physiologists and others of their kind.

---

[1] See, for example, some of the papers reprinted in Marx (1951).

When I compare psychologists with workers in other branches of science, and with my philosophical colleagues, I am forced to say that psychologists in general appear to suffer from methodological hypochondria, and from some personal insecurity about their intellectual standing and professional status in the community. I suspect that *one* of the sources of their hypochondria and insecurity is to be found in the fact that they have not yet managed to make inroads into the two problems of unity and vulnerability that I have picked out for discussion. If we are to accept the suggestions of psychotherapy, it is generally advisable for us to face up realistically to our worries and limitations. I hope that the brisk walk on which I have taken you this morning may do something to help us to face up to one source of our personal uneasiness, and something to show us what may be involved in removing it.

REFERENCES

ANDREWS, T. G. (ed.) (1948). *Methods of Psychology.* New York: Wiley; London: Chapman and Hall.
BROAD, C. D. (1933). *Examination of McTaggart's Philosophy,* vol. I. Cambridge University Press.
DEUTSCH, J. A. (1954). A machine with insight. *Quart. J. Exp. Psychol.* VI, 6–11.
GUTHRIE, E. R. (1935). *The Psychology of Learning.* New York and London: Harper.
HEBB, D. O. (1949). *Organization of Behaviour.* New York: Wiley; London: Chapman and Hall.
HULL, C. L. (1943). *Principles of Behaviour.* New York and London: Appleton.
MARX, M. H. (ed.) (1951). *Psychological Theory.* New York: Macmillan.
SKINNER, B. F. (1953). *Science and Human Behaviour.* New York: Macmillan.
STEVENS, S. S. (ed.) (1951). *Handbook of Experimental Psychology.* New York: Wiley; London: Chapman and Hall.
WITTGENSTEIN, L. (1953). *Philosophical Investigations.* Oxford: Blackwell.

(*Manuscript received* 30 *August* 1954)

THE IDEA OF INDEPENDENCE

BY JOHN COHEN AND C. E. M. HANSEL

Department of Psychology, University of Manchester

I. *Introduction* (p. 178). II. *Prediction of a binary event on the basis of previous outcomes* (pp. 178–181). III. *Effect of age on the prediction of outcomes of binary events* (pp. 182–183). IV. *Prediction of a binary event presented visually* (pp. 183–184). V. *Prediction of successive binary events in a series with knowledge of results after each choice* (pp. 184–185). VI. *Factors influencing the prediction of a binary independent event* (pp. 185–187). VII. *Conclusions and discussion of results* (pp. 187–189). VIII. *Appendix* (pp. 189–190). *Reference* (p. 190).

I. INTRODUCTION

We have been intrigued by the fact that the idea of an independent event is a very difficult one to grasp. Many people feel almost compelled to believe that after a run of successes failure is more likely to occur, or the other way round. If, for example, a coin has repeatedly turned up head, the chance of its continuing to do so is believed to get less with each further toss. Even those familiar with notions of statistical independence experience this difficulty. People seem to find it hard to consider an event as separate and detached from a series of similar events in which it occurs, although it may be perfectly evident that the outcome of the event in question can be in no way affected by the outcomes of the other events. The series of events is treated as one continuous single event. The great unlikelihood of getting, say, a run of 10 heads is attached to the possibility of getting a head on the tenth toss, after a run of 9 heads. Similarly, the extreme unlikelihood of not getting a single head in a sequence of 10 tosses is attached to the tenth toss after a run of 9 tails. This sort of reasoning is applied to all kinds of comparable situations in everyday life.

There are countless events in everyday life that can happen in one of two ways. For example, it may rain or not rain on a given day. A child born may be male or female. A candidate may pass or fail an examination. A paper submitted to the editor of a psychological journal may be accepted or rejected. Such events often occur in a sequence. There might be a succession of similar outcomes (e.g. a series of male births) or varying outcomes. In the present study of subjective probability we shall explore the way children at different ages predict the outcome of an event of this character when they are given some information about previous outcomes.

II. PREDICTION OF A BINARY EVENT ON THE BASIS OF PREVIOUS OUTCOMES

In this experiment we are concerned with the prediction of an event that can happen in one of two ways: a bead drawn from a given beaker can be either blue or yellow. The subjects are first given different sequences of previous outcomes and on the basis of these they make their predictions. The sequences are varied in order to discover the effect of particular sequences on the prediction.

The experimenter says to the subjects:

> I have here 12 beakers in a row. Each beaker contains blue and yellow beads. The proportions of blue and yellow beads vary from beaker to beaker. I am going to take out 4 beads, one at a time, from each beaker. I will show you the first three beads as they are drawn and then you will guess the colour of the fourth.

The colour of the fourth bead was not revealed.

The experiment was designed such that in any arrangement of 3 beads drawn from a beaker there were always 2 of one colour and 1 of the other in the six possible arrangements. The order of drawing the beads from the beakers was such that there were 2 beads of one and 1 bead of the other colour in the six possible arrangements. Each arrangement was repeated twice, the whole series being given in random order, as below:

Y B B	B B Y	B B Y
Y B Y	B Y B	Y B B
B Y B	Y Y B	Y Y B
B Y Y	B Y Y	Y B Y

The beads were drawn from each beaker in the order reading left to right.

We shall call the colour represented by the two beads the *preponderant* colour, and the colour represented by the single bead the *non-preponderant* colour. The terms preponderant and non-preponderant will be abbreviated as P and $\overline{P}$.

There are three different kinds of arrangement, disregarding colour. The preponderant and non-preponderant events are symbolized as 1, 0 respectively:

Y B B or B Y Y	0 1 1	B Y B or Y B Y	1 0 1	B B Y or Y Y B	1 1 0

Each of the three different kinds of arrangement may have either blue or yellow as the preponderant colour. By combining the arrangements in the manner shown, we remove any possible effect of colour preference. This method of combination enables us to assemble four choices made by each subject for each kind of arrangement. There were 93 subjects yielding 372 choices in all for each kind of arrangement (0 1 1; 1 0 1; 1 1 0). The number and proportion of choices for the preponderant and non-preponderant colour are shown, for each kind of arrangement, in Table 1.

Table 1. *Numbers and proportions of choices of preponderant and non-preponderant colours*

($N = 93$. Age 10 +)

	$\overline{P}$ followed by P 0 1 1		P in first and third positions 1 0 1		P followed by $\overline{P}$ 1 1 0		Total	
	Y B B or B Y Y		B Y B or Y B Y		B B Y or Y Y B			
Preponderant colour (P)	104	0·27	112	0·30	146	0·39	362	0·32
Non-preponderant colour ($\overline{P}$)	268	0·73	260	0·70	226	0·61	754	0·68
Total	372	1·00	372	1·00	372	1·00	1116	1·00

It may be said that the values shown in Table 1 have been obtained from twelve successive trials with different beakers and that later guesses are affected by earlier trials. We find, however, that a value of 0·27 (for the non-preponderant colour) is obtained in the very first trial of 1 1 0.

The first point to note is that the non-preponderant colour is in each case favoured in the fourth choice. This fourth choice seems to be made in accordance with a principle of compensation or balance which equalizes the possibilities of the two outcomes. It is not made independently of previous occurrences, otherwise there would be no difference in the number of choices of preponderant and non-preponderant colour. Nor is it in accord with a statistical prediction based on sampling from a large population. Two explanations suggest themselves. First, the choice of $\overline{P}$ might presuppose an implicit assumption that an initial store of blue and yellow beads of equal numbers in the beakers has been unevenly depleted. If, for example, a child implicitly assumed that there were at the start 3 blue and 3 yellow beads and he then saw 2 blue and 1 yellow drawn from the beaker, he would presumably conclude that 1 blue and 2 yellow remain and therefore he would prefer to predict yellow. Secondly, and this does not exclude the first explanation, the tendency to favour $\overline{P}$ as the fourth choice may be due primarily to the asymmetrical character of the arrangement of beads already drawn. We have indeed found in previous experiments (Cohen & Hansel, 1955) a tendency to prefer symmetrical arrangements of 4 beads under conditions when there is no question of depleting a store.

Another point to note is that the order of appearance of the colours affects the fourth choice. The more remote the appearance of the non-preponderant colour the more frequently it is chosen. This effect is rather like the one mentioned above as frequently noticed in everyday life in which a person's expectation of heads increases with more appearances of tails.

From earlier experiments (Cohen & Hansel, 1955) we know that subjects tend to assume equal proportions when the proportions are stated to be unknown. If they made such an assumption here, it would not be contradicted by what they saw emerging from the beaker, for if three beads are drawn, a 2:1 ratio of colours is the nearest to a 1:1 ratio that is possible to occur. The fact that they choose $\overline{P}$ suggests that they make this assumption and their choice of $\overline{P}$ enables them to maintain it.

We now turn to experiments in which the results of previous events were communicated to the subjects verbally, and at the same time illustrated visually on a blackboard. The kind of binary event now ranges from independent occurrences at one end to dependent occurrences at the other. The events about which a prediction has to be made are such as come within the experience of children. They include independent occurrences, such as coin tosses and the sex of the next baby to be born in a given town, and dependent occurrences, such as the state of the weather on a given day and which one of two uncles has sent a present. These instances of dependence were selected because it was thought the outcomes would readily be predicted on the basis of previous experience.

The procedure was to say to the subjects:

> A boy has two uncles, Uncle George and Uncle John. In 1951 he received a present from Uncle John but not from Uncle George. In 1952, he received a present from Uncle George but not from Uncle John. In 1953 he also received a present from Uncle George but not from Uncle John. In 1954 he has received a present and on the label there are the words 'from Uncle'. Which uncle do you think sent the present?

The appropriate wording was modified to suit the other cases.

The proportions in Table 2 are based on samples of 44. In the two cases marked by an asterisk, $N = 88$. All the proportions shown in the table are significant at the 0·01 level. The choice of the non-preponderant event in these experiments is more favoured than in the experiments with beads. This may be due to the tendency to complete or maintain the pattern that appears in the visual illustration of previous occurrences. For example, the problem $\begin{matrix}1&0\\0&1\\1&0\end{matrix}$ is presented on the blackboard as follows:

	Boy	Girl
1951	√	—
1952	—	√
1953	√	—
1954?		

Moreover, visual illustration also strengthens the tendency to alternate, as may be seen by comparing the middle row of Table 2 with the first and third rows.

Table 2. *Proportions of choices between alternatives in* 1954

			Winning a toss		Sex of baby		Rain or fine		Sender of present		Mean values	
Year	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$
1951	0	1	0·15	0·85	0·12	0·88	0·12	0·88	0·25	0·75	0·16	0·84
1952	1	0										
1953	1	0										
1951	1	0	0·07	0·93	0·10	0·90	0·07	0·93	0·07	0·93*	0·07	0·93
1952	0	1										
1953	1	0										
1951	1	0	0·12	0·88*	0·32	0·68	0·20	0·80	0·23	0·77	0·22	0·78
1952	1	0										
1953	0	1										

* Means of the proportions made when choices are presented in two orders, e.g. $\begin{matrix}1&0\\0&1\\1&0\end{matrix}$ and $\begin{matrix}0&1\\1&0\\0&1\end{matrix}$, given to two groups of 44 children.

The desire to continue the visual pattern so as to keep or make it balanced may enhance the tendency to choose the non-preponderant value as compared with the choices in the previous experiment with beads. This suggestion is supported by the results obtained when a similar type of problem—in this instance relating to true events—was given without illustration by visual pattern. The extent to which the non-preponderant was favoured in these circumstances was of the order found in the experiments with beads. The problem was of the form $\begin{matrix}0&1\\1&0\\1&0\end{matrix}$ and presented as follows to 89 ten year-old children:

Oxford won the boat race this year and last year. Cambridge won two years ago. Who do you think will win next year?

The choices were 0·38 in favour of Oxford, 0·62 in favour of Cambridge, i.e. the non-preponderant event ($\chi^2=4\cdot9$, D.F.$=1$, $p<0\cdot05$).

The alternation or choice of the non-preponderant event also appears significantly, under similar conditions, in the cases $\begin{matrix}0&1\\1&0\end{matrix}$ (0·36, 0·64) and $\begin{matrix}1&0\\1&0\end{matrix}$ (0·41, 0·59), but in less marked fashion. In the cases $\begin{matrix}1&0\end{matrix}$ and $\begin{matrix}1&0\\1&0\\1&0\end{matrix}$ there appears to be no general tendency to alternate or to favour the non-preponderant event.

III. EFFECT OF AGE ON THE PREDICTION OF OUTCOMES OF BINARY EVENTS

We now consider in relation to one kind of independent event, coin tossing, the predictions made at three ages, 10, 12 and 14 years. These are shown in Table 3. A small number of subjects did not choose P or $\bar{P}$ but replied 'either'. These have been divided between the two choices.

Table 3. *Proportions of choices of alternatives made at different ages*

($10+(N=50)$, $12+(N=64)$, $14+(N=57)$: winning a toss)

Year	P	$\bar{P}$	10+		12+		14+	
			P	$\bar{P}$	P	$\bar{P}$	P	$\bar{P}$
1951	0	1	0·15	0·85	0·19	0·81	0·29	0·71
1952	1	0						
1953	1	0						
1951	1	0	0·07	0·93	0·23	0·77	0·38	0·62
1952	0	1						
1953	1	0						
1951	1	0	0·12	0·88	0·34	0·66	0·38	0·62
1952	1	0						
1953	0	1						

Of the nine sets of values in this table, seven deviate significantly from equal proportions at least at the 0·05 level of significance. Two values, 0·38, 0·62 (at the age 14+) are barely significant. They become significant however, if we exclude subjects giving judgements of 'either'. There were 9 in each case, the number remaining out of 57 is therefore 48, of which 17 chose P and 31 chose $\bar{P}$ ($\chi^2 = 4{\cdot}3$, D.F. $=1$, $p < 0{\cdot}05$). It follows that in the circumstances of this experiment there is a decided tendency in the members of the group to predict what has happened less often in the past three events as more likely to happen in the future.

It is clear from this table that between the years 10 to 15 the tendency to favour the non-preponderant event weakens. In the three cases combined there are 150 choices made by the 10+ group, 192 by the 12+ group and 171 by the 14+ group. The numbers of P and $\bar{P}$ choices given by the three groups respectively are:

	10+	12+	14+
P	17	49	60
$\bar{P}$	133	143	111
Total	150	192	171

We can compare the proportions in the 10+ and 12+ groups and in the 12+ and 14+ groups giving the different choices. $\chi^2 = 10{\cdot}9$, D.F. $=1$, $p < 0{\cdot}001$ in the first instance; and 3·9, D.F. $=1$, $p < 0{\cdot}05$ in the second instance. The difference between the 10+ and 14+ is *a fortiori* significant.

This decline with age in the preference for the non-preponderant value in the fourth choice has its parallel in the increased tendency with age to give an 'either' judgement, a tendency which amounts to a recognition of a principle of independence. The ten-year-olds did not give a single 'either' judgement, the twelve-year-olds gave 15 (out of 192 choices) and the fourteen-year-olds gave 26 (out of 171 choices) in reply to the three questions.

In all three age groups the tendency to alternate or to choose $\overline{P}$ appears in the cases $\begin{smallmatrix}0&1\\1&0\end{smallmatrix}$ and $\begin{smallmatrix}1&0\\1&0\end{smallmatrix}$ but not in $1\;0$ and $\begin{smallmatrix}1&0\\1&0\\1&0\end{smallmatrix}$. Similar results were obtained with the other kinds of event shown in Table 2.

IV. Prediction of a binary event presented visually[1]

In order to extend the range of conditions under which forms of subjective probability may be investigated and also to standardize a procedure suitable for subjects at all ages we constructed an apparatus consisting of a display board with 16 pairs of lights, as shown in Fig. 1.

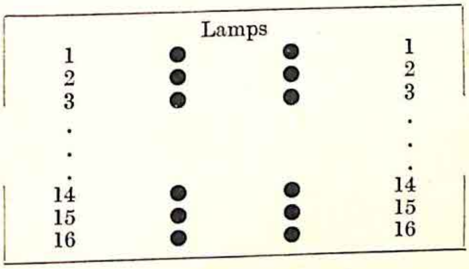


Fig. 1. Display board.

In the experiment, one of each of the first 10 pairs was illuminated, 9 being on one side and 1 on the other. The position of the single light varied (in random order) from position 1 to position 10 in different trials. The preponderant side bears 9 lights and the non-preponderant side a single light. In each set of 10 predictions, made by each subject, the preponderant side was on the right on five occasions and on five occasions on the left.

After seeing the lights 1 to 10 illuminated in rapid succession and remaining lit, the subject had to guess on which side the eleventh light would appear. In fact the eleventh light was never illuminated. The subjects were studied individually. We explained to them that only one of each pair of lights would go on. They indicated their choice by pointing to the small lamp in the eleventh row which they expected to be illuminated next.

The subjects in this experiment were 16 six-year-old and 8 seven-year-old children. Their predictions are shown in Table 4.

Considering first the possible effect of the position in the series of the single light, we did not find, what might have been expected, that the tendency to predict the eleventh light on $\overline{P}$ increased as its position occurred earlier in the series of 10 lights. This may be due to the fact that the display, when all 10 lights were lit, resembled a stable visual pattern and that it was in terms of visual patterns of lights that predictions were made and not in terms of remoteness of occurrence in the past.

Taking next the question whether the non-preponderant side is favoured at all in this experiment, at first sight it does not appear so to any marked extent in the six-year-olds. We found, however, that both the six-year-old and the seven-year-old children tended to alternate their choices of right and left throughout the 10 successive predictions. There were 147 alternating as against 69 non-alternating predictions. In other words, their choices were to a large extent determined by what they had chosen in the previous situation, not by the relations of P and $\overline{P}$ confronting them. On reclassifying the choices into alternating and non-alternating respectively (as in Table 4), we find that when they do not alternate, they do tend to favour the non-preponderant side.

---

[1] We are grateful to Mr R. Church for designing and constructing the display board and for operating it during the experiments.

Table 4. *Prediction of eleventh light on preponderant or non-preponderant side*

($N = 16$ (6+), 8 (7+).)

	Alternation			Non-alternation			
Age	P	$\overline{P}$	Total	P	$\overline{P}$	Total	Total
6+	58	50	108	13	23	36	144
7+	21	18	39	12	21	33	72
Total	79	68	147	25	44	69	216

On the null hypothesis we should expect an equal number for P and $\overline{P}$ in the non-alternating choices. In fact the number favouring $\overline{P}$ in the combined group of 24 children is significantly large ($\chi^2 = 5{\cdot}3$, D.F. $= 1$, $p < 0{\cdot}03$). The tendency to alternate, i.e. to be influenced by choice in a preceding situation, seems to be considerably weaker in the seven-year-old children. These seem more capable of making a choice on the basis of the display confronting them. In so far as they do not alternate the behaviour of the two age groups in relation to P and $\overline{P}$ is much the same.

V. PREDICTION OF SUCCESSIVE BINARY EVENTS IN A SERIES WITH KNOWLEDGE OF RESULTS AFTER EACH CHOICE

The problem here is to identify the influences affecting the prediction of a binary outcome in a series. Among these influences are: (i) the immediately preceding event, i.e. the actual side on which the previous light appeared; (ii) the number of lights which have already appeared on each side; (iii) the immediately preceding choice, i.e. the side *predicted* for the previous light; and (iv) success or failure in the previous prediction.

The apparatus consisted of the display board with 16 pairs of lights already described. This time before any one of a pair of lights was switched on, the subject had to guess which member of the pair it would be. Then one of the first pair of lights was switched on according to a pre-arranged scheme and remained on. The same procedure was followed with each pair of lights in succession from the second to the sixteenth pair. In one form of the experiment the position of the lights was randomized in a 1:1 ratio so that there finally appeared 8 lights on each side in random order; and in the other form of the experiment the position of the lights was randomized in a 3:1 ratio so that finally 4 lights appeared on one side of the display board and 12 lights on the other. There were two groups of subjects, 46 aged 10+ and 16 aged 6–8 years.

From the successive choices in the experiment with a 1:1 ratio in which 8 lights eventually appeared in different orders on each side of the display board the four influences can be assessed.

(i) *Effect of the side on which the previous light appeared*

In both age groups there is a significant tendency to alternate from the side of the previous light. If the previous light is on the right, the choice tends to be on the left and vice versa. χ^2 for the 10+ group is 8·0 (D.F. $= 1$, $p < 0{\cdot}01$) and for the 6–8 group, 20·2 (D.F. $= 1$, $p < 0{\cdot}001$).

(ii) *Effect of number of lights which have already appeared on each side*

Here again at both ages there is a marked tendency to predict the next light on the side with fewer lights. This is a preference for $\overline{P}$. χ^2 for the older group is 32·9 (D.F. $=2$, $p<0{\cdot}001$), and for the younger group 7·8 (D.F. $=2$, $p<0{\cdot}05$). From the proportions making the different choices at both ages it is evident that the older children are much more affected by the total number of lights that have already appeared than the younger children.

(iii) *Effect of side chosen for previous light*

In neither of the two age groups is the choice of any side influenced by the side previously chosen. This means that the children are not affected by what they have previously predicted but by the previous outcome.

(iv) *Effect of success or failure in the preceding choice*

Here there is a marked age difference. The younger children alternate after success but not after failure. $\chi^2=21{\cdot}1$ (D.F. $=1$, $p<0{\cdot}001$). The older children show no tendency to alternate merely because they have either succeeded or failed in their preceding guess.

In the second form of this experiment in which the final ratio of lights was 3:1 we were interested to see whether, when the disparity between the number of lights on either side increased, the children tended to guess the next light as due to appear on the side with more lights. In the early predictions the choices are not unlike those made when the ratio is 1:1. With the 3:1 display, however, a tendency gradually appears to prefer the preponderant side. In the last four choices with the 1:1 display there is no significant tendency to prefer one or the other side, whereas with the 3:1 display there is a significant tendency to prefer P ($\chi^2=8{\cdot}5$, D.F. $=1$, $p<0{\cdot}01$).

VI. FACTORS INFLUENCING THE PREDICTION OF A BINARY INDEPENDENT EVENT

The following experiment is an attempt to investigate the everyday phenomenon mentioned in the first paragraph that one outcome of a binary event is felt to be more likely as the run of the other outcome is prolonged. The subjects here are girls aged 12–15 years. The two age groups 12+ ($N=64$) and 14+ ($N=57$), each comprised two subgroups of higher and lower ability, as judged by the school authorities.

The kind of activity giving the independent event, i.e. coin-tossing, selected for this experiment, has the advantage of resembling a familiar situation. Furthermore, it can be thought of in two ways: (i) in terms of runs of head or tail, (ii) in terms of runs of success or failure of the persons tossing the coin. Two forms of question were used:

1. A girl is tossing a penny. The coin has turned up head on all of the last x tosses. What do you think will turn up the next time it is tossed?

In an alternative form of the above question, 'tail' was substituted for 'head'.

2. Joan and Mary, two captains of school teams, are about to toss a coin before a match. Joan has won and Mary has lost on the last x guesses. Who do you think is more likely to win next time?

In an alternative form of this question the names Joan and Mary were interchanged. In the space marked x a different value from 1 to 10 was placed, a single value being included in the question presented to each subject. Each subject thus made one guess in reply to one of the four types of question. After recording their guess, the subjects were asked to explain why they had given that particular answer.

When we classify the guesses according to length of run of one outcome, we find that 54 out of 121 favour a change in victor or outcome. This number is not, however, significantly different from the number (45) favouring a continuance of the run. Nor does the length of run of one outcome seem to have had any consistent effect on the guess. The proportions favouring each of the two possible outcomes are shown in Table 5.

Table 5. *Proportions favouring continuance or change of run*

Length of run	Continuance of run	Change of outcome	Total
1	0·55	0·45	1·00
2	0·40	0·60	1·00
3	0·45	0·55	1·00
4	0·33	0·67	1·00
5	0·44	0·56	1·00
6	0·55	0·45	1·00
7	0·33	0·67	1·00
8	0·50	0·50	1·00
9	0·43	0·57	1·00
10	0·63	0·37	1·00

In both age groups the proportion of abler subjects giving an 'either' judgement is significantly greater than the proportion choosing one or other outcome. In this respect the abler younger ones resembled the abler older ones more than they resembled their contemporaries.

The fact that we did not obtain a significantly clear-cut tendency in this experiment may be due to the conditions of the experiment: the subjects were not participants in a real situation, they merely answered questions about a hypothetical situation. Other techniques (which we intend to employ) may reveal the alleged effect more definitely.

Our present results suggest that the phenomenon is interwoven with other effects which tend to mask it. The reasons given by the subject for their choices enable us to understand the way this happens. At least 48 subjects stated that they believed it was time for the non-successful outcome to occur, or words to that effect. But in many instances they also adduced other reasons, such as that the tosser was skilful and so the run of successes would continue. Indeed there are a large variety of reasons given and these are often of a highly composite character. It is clear that the decision or choice is far from being a simple one in the majority of instances, and varies considerably from subject to subject. The explanations have been grouped together under the headings: Dependence (including Favouring change, Luck, Determinate, Series and Magic) and Independence.

Some subjects gave more than one reason, so the total number of reasons shown in Table 6 exceeds the size of the sample.

The main features that emerge from the classification of reasons are (i) the relatively high proportion of the younger age-group, as compared with the older, giving 'luck' as the reason for their choice, $\chi^2 = 4{\cdot}1$, D.F. = 1, $p < 0{\cdot}05$, and (ii) the relatively small proportion of the younger group who say that the outcome cannot be known, $\chi^2 = 6{\cdot}8$, D.F. = 1, $p < 0{\cdot}01$.

Perhaps the simplest interpretation of these results is in terms of treatment of the unknown outcome as dependent or independent of previous outcomes. The proportions of these two types in the total explanations are about 85 and 15% in the younger group and 73 and 27% in the older group. Those giving explanations in terms of dependence offer two antithetical elements. Some children treat the dependence so as to lead to continuance of the same outcome. Other children treat the dependence so as to lead to a change of outcome. The former group use the data as providing information about the personal characteristics of the tosser such as skill, determination or luck or about the characteristics of the coin, and they extrapolate this to the next throw. The latter group consider the unknown outcome as determined by previous outcomes but not because these previous outcomes reveal the characteristics of the tosser or the nature of the coin. The mere fact that an outcome has occurred once or more times or that one person has had a run of luck means for them that there will be a change of outcome. In the simplest form this is expressed in the notion of alternation, e.g. 'I believe Mary will win next time because it goes alternately'.

Table 6. *Frequency of reasons given for prediction of outcome*

	Age	
	12–13*	14–15†
Dependence		
Favouring change	24	24
Luck	23	11
Determinate	16	18
Series	6	8
Magic	3	4
Total	72	65
Independence	13	24
Total	85	89

* $N = 32 + 32 = 64$.
† $N = 29 + 28 = 57$.

VII. CONCLUSIONS AND DISCUSSION OF RESULTS

The conclusions of the various experiments described above are briefly summarized:

(i) (*a*) After seeing 3 beads each of which may be one of two colours drawn from a beaker containing unknown proportions of blue and yellow beads, the non-preponderant colour is favoured in the fourth choice.

(*b*) The order of appearance of the colours affects the fourth choice, the more remote the appearance of $\overline{P}$ the more frequently it is chosen.

(ii) In comparable experiments with varying kinds of event the previous outcomes of which are visually illustrated, $\overline{P}$ is even more strongly favoured than in the foregoing experiments with beads.

(iii) $\overline{P}$ is more favoured in independent kinds of event such as coin-tossing and less favoured (as compared with coin-tossing) in events where the outcome appears strongly dependent on previous outcomes (e.g. 'sender of present').

(iv) Between the ages of 10 and 15 years the tendency to favour $\overline{P}$ weakens and the tendency to give 'either' judgements increases.

(v) (*a*) In experiments with children aged 6+ and 7+, using a display board with 9 lights on one side and the tenth light occupying any position from 1 to 10, we found a marked tendency to alternate choices from right to left in the six-year-olds but scarcely any such tendency in the seven-year-olds.

(*b*) In so far as they do not alternate they tend to favour $\overline{P}$.

(vi) In experiments with 6+, 7+ and 10+, using the display with a 1:1 ratio of lights appearing at random, we found the following tendencies:

(*a*) To change from the side of the previous light.

(*b*) To change from the side with more lights to the side with fewer lights.

(*c*) An age difference in the effect of success or failure in the preceding choice. The 6–8 group alternate after success but not after failure. The older children show no tendency to alternate merely because they have either succeeded or failed.

(vii) In experiments similar to those in (vi) above, but employing a 3:1 ratio of lights, the early predictions resembled those obtained under 1:1 conditions. The later predictions tended in the direction of the side with more lights.

(viii) (*a*) In experiments with groups of children 12+ and 14+ on prediction of an unknown binary outcome after varying length of run of one outcome, no striking increase in the tendency to change was found as the length of run was prolonged. This was interpreted, on the basis of subjects' protocols, as due to two antithetical ideas of dependence in this situation. Sometimes the antithesis was present in the explanation of one and the same person.

(*b*) Fourteen-year-olds gave more 'either' judgements than the twelve-year-olds, and the abler children at both age levels gave more 'either' judgements than the relatively less able.

(*c*) The frequency of explanations of choice in terms of 'luck' declined and in terms of indeterminateness increased in the 14+ as compared with the 12+ group.

We may now attempt to trace the development of the idea of independence in so far as it emerges in these experiments. At the youngest age which we have investigated, 6 to 7 years, the most marked feature is a tendency to alternate from the previous choice in situations of binary outcome. To the extent that they do not alternate their choices they are apt to choose the non-preponderant outcome. From their remarks it seems that they regard this as fair. They are also influenced by whether they have succeeded or failed in the last attempt: after success they alternate, after failure they do not. To a relatively slight extent they also tend to alternate from the side of the previous outcome and from the side with more outcomes. In short the six-year-olds show a powerful tendency to alternate as a result of a variety of specific influences.

Children a year older are much less apt to alternate their choice. Like some of the six-year-olds but to a more marked degree, the seven-year-olds tend to change from the side of the previous outcome and from the side with more outcomes to the side with less outcomes. But unlike the younger ones they do not tend to alternate from previous choice and mere success or failure in the preceding choice does not influence them. They are much more under the influence of the display confronting them.

We have found that most children, at least till the age of 10+, do not initially distinguish between an event the outcome of which is in reality independent of previous outcomes, such as tossing a coin, and an event with a varying degree of dependence on previous outcomes, such as fine or not-fine weather in a given place on a certain day.

Judgements on these two types of situation are similar for short runs of, say, eight events. $\overline{P}$ is favoured even when knowledge of results shows that one event arises three times as often as the other. Children may treat such short series of independent and dependent events without distinction as single unit events, hence favouring $\overline{P}$. This effect is very marked in series consisting of five or six events.

In longer series, however, they do seem to make the distinction. When the event is of the coin-tossing (i.e. independent) kind, the children seem to predict in longer series in much the same way as they do in shorter series. When, however, as with the display board, the event in the early part of the series may be equally regarded as either of one kind or of the other, the predictions in the later part of the series (after approximately eight outcomes) follow the preponderant outcome. This happens when the ratio of outcomes is 3:1 and would no doubt be even more marked with higher ratios.[1]

An idea of independence manifested in a judgement of 'either' does not appear in our experiments at the age 10+. It first emerges at the age 12+. From then onwards the tendency to favour the non-preponderant outcome weakens, together with a strengthening of the inclination to give 'either' judgements.

Decisive changes in the idea of independence as held by some subjects take place between the ages of 12+ and 15+. The idea of luck has a variety of forms and meanings at this period. Gradually it appears to merge in the notion of chance as a concept of independence. For instance a child says: 'It is just luck.' There are two opposing ideas of dependence at this time sometimes held by the same subject. One of these ideas of dependence favours continuance of a series of similar outcomes. The other favours a change to the other binary outcome as the series proceeds. The idea of independence is beginning to emerge at the age 12+ and is clearly and forcefully held by some subjects at the age of 14+. Even at this age, however, the vast majority entertain ideas of dependence; the idea of luck is still potent, and ideas of magic are not lacking.

VIII. APPENDIX

Some typical explanations of the reasons for the choice of the outcome in section VI are given below:

Favouring change

'Because if Mary has won 7 times she is not likely to win again.'

'Mary (will win) because Joan won all the other times and it is about time Mary won.'

'I think Mary will win because on the Law of Averages it is Mary's turn to win.'

'Joan will probably win the toss because usually the one who lost the time before will win the next time.'

'Because usually the one who has been losing a lot wins. But really it is just luck.'

Luck

'One girl has won 4 times but her luck has run out so the second girl wins and the first girl loses because tails has not won.' (This also favours change.)

'Mary seems to be the lucky one of the two and she won first time.'

'Joan seems to have all the luck so I'm sure she will win again.'

'I think it would be tails because most people think it is the luckier of the two.'

---

[1] If all the lights were on one side (i.e. in the ratio 1:0) we should have a simple learning situation.

(v) (*a*) In experiments with children aged 6+ and 7+, using a display board with 9 lights on one side and the tenth light occupying any position from 1 to 10, we found a marked tendency to alternate choices from right to left in the six-year-olds but scarcely any such tendency in the seven-year-olds.

(*b*) In so far as they do not alternate they tend to favour $\overline{P}$.

(vi) In experiments with 6+, 7+ and 10+, using the display with a 1:1 ratio of lights appearing at random, we found the following tendencies:

(*a*) To change from the side of the previous light.

(*b*) To change from the side with more lights to the side with fewer lights.

(*c*) An age difference in the effect of success or failure in the preceding choice. The 6–8 group alternate after success but not after failure. The older children show no tendency to alternate merely because they have either succeeded or failed.

(vii) In experiments similar to those in (vi) above, but employing a 3:1 ratio of lights, the early predictions resembled those obtained under 1:1 conditions. The later predictions tended in the direction of the side with more lights.

(viii) (*a*) In experiments with groups of children 12+ and 14+ on prediction of an unknown binary outcome after varying length of run of one outcome, no striking increase in the tendency to change was found as the length of run was prolonged. This was interpreted, on the basis of subjects' protocols, as due to two antithetical ideas of dependence in this situation. Sometimes the antithesis was present in the explanation of one and the same person.

(*b*) Fourteen-year-olds gave more 'either' judgements than the twelve-year-olds, and the abler children at both age levels gave more 'either' judgements than the relatively less able.

(*c*) The frequency of explanations of choice in terms of 'luck' declined and in terms of indeterminateness increased in the 14+ as compared with the 12+ group.

We may now attempt to trace the development of the idea of independence in so far as it emerges in these experiments. At the youngest age which we have investigated, 6 to 7 years, the most marked feature is a tendency to alternate from the previous choice in situations of binary outcome. To the extent that they do not alternate their choices they are apt to choose the non-preponderant outcome. From their remarks it seems that they regard this as fair. They are also influenced by whether they have succeeded or failed in the last attempt: after success they alternate, after failure they do not. To a relatively slight extent they also tend to alternate from the side of the previous outcome and from the side with more outcomes. In short the six-year-olds show a powerful tendency to alternate as a result of a variety of specific influences.

Children a year older are much less apt to alternate their choice. Like some of the six-year-olds but to a more marked degree, the seven-year-olds tend to change from the side of the previous outcome and from the side with more outcomes to the side with less outcomes. But unlike the younger ones they do not tend to alternate from previous choice and mere success or failure in the preceding choice does not influence them. They are much more under the influence of the display confronting them.

We have found that most children, at least till the age of 10+, do not initially distinguish between an event the outcome of which is in reality independent of previous outcomes, such as tossing a coin, and an event with a varying degree of dependence on previous outcomes, such as fine or not-fine weather in a given place on a certain day.

Judgements on these two types of situation are similar for short runs of, say, eight events. $\overline{P}$ is favoured even when knowledge of results shows that one event arises three times as often as the other. Children may treat such short series of independent and dependent events without distinction as single unit events, hence favouring $\overline{P}$. This effect is very marked in series consisting of five or six events.

In longer series, however, they do seem to make the distinction. When the event is of the coin-tossing (i.e. independent) kind, the children seem to predict in longer series in much the same way as they do in shorter series. When, however, as with the display board, the event in the early part of the series may be equally regarded as either of one kind or of the other, the predictions in the later part of the series (after approximately eight outcomes) follow the preponderant outcome. This happens when the ratio of outcomes is 3:1 and would no doubt be even more marked with higher ratios.[1]

An idea of independence manifested in a judgement of 'either' does not appear in our experiments at the age 10+. It first emerges at the age 12+. From then onwards the tendency to favour the non-preponderant outcome weakens, together with a strengthening of the inclination to give 'either' judgements.

Decisive changes in the idea of independence as held by some subjects take place between the ages of 12+ and 15+. The idea of luck has a variety of forms and meanings at this period. Gradually it appears to merge in the notion of chance as a concept of independence. For instance a child says: 'It is just luck.' There are two opposing ideas of dependence at this time sometimes held by the same subject. One of these ideas of dependence favours continuance of a series of similar outcomes. The other favours a change to the other binary outcome as the series proceeds. The idea of independence is beginning to emerge at the age 12+ and is clearly and forcefully held by some subjects at the age of 14+. Even at this age, however, the vast majority entertain ideas of dependence; the idea of luck is still potent, and ideas of magic are not lacking.

VIII. APPENDIX

Some typical explanations of the reasons for the choice of the outcome in section VI are given below:

Favouring change

'Because if Mary has won 7 times she is not likely to win again.'

'Mary (will win) because Joan won all the other times and it is about time Mary won.'

'I think Mary will win because on the Law of Averages it is Mary's turn to win.'

'Joan will probably win the toss because usually the one who lost the time before will win the next time.'

'Because usually the one who has been losing a lot wins. But really it is just luck.'

Luck

'One girl has won 4 times but her luck has run out so the second girl wins and the first girl loses because tails has not won.' (This also favours change.)

'Mary seems to be the lucky one of the two and she won first time.'

'Joan seems to have all the luck so I'm sure she will win again.'

'I think it would be tails because most people think it is the luckier of the two.'

---

[1] If all the lights were on one side (i.e. in the ratio 1:0) we should have a simple learning situation.

Determinate

'Because Mary most likely can only throw a certain way and Joan knows what it usually comes to.'

'Mary won the last two so she must be the best and if she is the best she will surely win the next one.'

'Because if Joan has won 6 already she is obviously better than Mary and so it is possible that she will know more the next time.'

'Because the wind might be causing the coin to drop on the side that it has been doing previously.'

'Joan (will win). Joan would be very determined to win again and if the same person tossed she would probably call what she called the last time which was heads. Mary would be very nervous and have given up hope of winning.'

Series

'Tails. Because it has turned up tails 3 times already and it is sure to turn up the same after three times turning up tails.'

'I think heads will turn up next time because it has for 5 times.'

'It will turn up Heads because if it has turned up the same ten times before it will probably do so again.'

'Mary is likely to win because the coin usually turns up six or more times the same, if it has turned up five.'

Magic

'Mary will think she will win and so probably will not win.'

'Heads. Because for some strange reason, the fifth toss always comes heads. I read this in a book, and tried it often, and it always works (book title, 1000 *Wonderful Things*).'

'If the girl chose tails I think it will be tails because there is a saying that "Tails, tails never fails". This saying is often very true.'

'Tails. Because 7 is my lucky number and it is the seventh time it has been tossed after 6 heads. So I think it will change to tails by the 7th time.' (This also involves the idea of luck.)

Indeterminate

'It depends on the way the coin is tossed up. Any one of them might win. Mary has won the guesses but this does not mean that she will win this time. She might very well but Joan could win this time. It is not possible to tell who will win.'

'Heads or tails. It's just up to the penny.'

'Because if a penny spins in the air it does not necessarily land on the same side every time.'

'Because a penny has two sides, a head and a tail, and no one can tell which end it is going to land (unless they cheat) so it could be either heads or tails.'

'Because when you toss a penny nobody, *nobody* can tell even the cleverest of people which it is going to be, because it might be either.'

'Either. It just depends on which side it lands after being tossed.'

REFERENCE

COHEN, J. & HANSEL, C. E. M. (1955). The idea of a distribution. *Brit. J. Psychol.* XLVI, 111–21.

(*Manuscript received* 1 *July* 1954)

PROBLEM-SOLVING AS DEPENDENT ON AVAILABILITY OF FUNCTIONS[1]

By PER SAUGSTAD

University of Oslo

I. *Introduction* (pp. 191–193). II. *Experimental* (pp. 193–196): (i) *Subjects* (p. 193); (ii) *The problem* (pp. 193–194); (iii) *Procedure* (pp. 194–195); (iv) *Results* (pp. 195–196). III. *Discussion* (pp. 196–197). IV. *Summary and conclusion* (pp. 197–198). *References* (p. 198).

I. INTRODUCTION

There exists, as yet, no satisfactory definition of what constitutes a 'problem'. Perhaps the nearest approach to such a definition is that of Humphrey (1951): 'A problem is a situation which for some reason appreciably holds up an organism in its efforts to reach a goal' (p. 312). However, whatever the nature of a problem, it is fairly certain that its solution involves the use of past experience and the combination of certain 'functions' —or to use Harlow's (1951) phrase, 'organized response patterns'—which, as part of the problem situation, are initially discrete. The purpose of this paper is to explore the possibility of analysing problem-solving in terms of 'functions', with particular emphasis on their availability to the subject at the time of attempting the solution and on their combination.

The present study arose out of a consideration of the work of Maier (1930, 1933, 1937, 1945) and is perhaps best introduced by a consideration of Maier's pendulum problem. The task was to construct two pendulums, each of which would swing over a designated point on the floor and make a chalk mark at that point. Various materials were available to the subjects: a large table, two long poles, two shorter poles, one table clamp, two burette clamps, two lengths of flexible wire, eight pieces of lead tubing and several pieces of chalk. The solution involved three steps: (1) constructing the pendulums by inserting a piece of chalk into a clamp tied to the end of a piece of wire, (2) making a pole long enough to reach the ceiling by clamping two poles together, and (3) supporting the two pendulums at either end of a third pole which was then wedged against the ceiling with the two joined poles. To ensure that his experimental group had the necessary experience to solve the problem (i.e. the necessary knowledge of the three steps involved), Maier had previously shown them (1) how to make a plumb line out of a burette clamp and a pencil, (2) how to combine two shorter poles into a longer one by clamping them together, and (3) how to wedge a piece of cloth against the wall of a door-way by pressing the side of a flat board against it and then keeping this in place by wedging another board horizontally between the flat board and the other wall of the door-way. The subjects in the group given this pre-problem experience did not solve the pendulum problem although, argued Maier, they had sufficient knowledge to do so. Weaver & Madden's (1949) repetition of this experiment gave essentially the same results.

Now, in asserting that the above subjects had all the experience necessary to the solution of the problem, Maier was rather rash. After all, a particular experience always involves a specific, concrete situation, and it is from a succession of such particular experiences that principles, concepts, universals (whatever we choose to call them) are derived. Maier demonstrated the function of certain objects in one definite situation, but to solve the problem, other objects had to be used in another situation. The pre-problem demonstrations would only assist the subjects in solving the pendulum problem if they were seen as illustrating principles of fairly wide applicability. Thus, the situation in which the cloth is wedged against the door-post by the flat board is not the same as that in which a pole is wedged against the ceiling. Even if the sight of the pole elicited the principle of wedging, there still remained the problem of finding a second flat surface (the ceiling). The reality of this remaining problem is obvious from the performance of a second experimental group to which Maier gave so-called 'direction' by demonstrating how the pendulum problem might be solved by nailing the wire to the ceiling. More members of this group solved the pendulum problem than of the other experimental group: thus, it was of definite assistance to have the possible function of the ceiling pointed out. If, then, subjects are furnished with an experience in order to ensure their having a given principle, it is necessary to ascertain whether or not this principle has, in fact, been grasped. The distinction between the particular and the general is one which Maier failed to take into account when discussing the sufficiency of pre-problem experiences.

---

[1] This study was conducted while the writer was Rockefeller fellow in Psychology at the University of Chicago. The article is based upon a dissertation submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy at the University of Chicago.

Another instance of inadequate handling of pre-problem experiences occurs in a later experiment by Maier (1945). One group of subjects were shown how to build a structure similar to one of the pendulums and then afterwards given the problem of constructing a hat rack. To solve this problem it was necessary to clamp two poles together and wedge them between the floor and the ceiling so that the clamp could function as a hook on which to hang a hat and a coat. Seventy-two per cent of the subjects solved this problem. A second group was present only during the construction of the hat rack, although no mention was made of its possible functions. When this group was asked to list all the possible functions which the construction might have, only 24 % mentioned its use as a hat rack. If I interpret Maier aright, he argues that, since these two groups had both seen the same construction, they had been exposed to equivalent stimulus situations. However this may be, it seems clear that the two groups received markedly different instructions and so faced the situation with different attitudes. The first group was confined by the instructions to the concept of a hat rack. The second group was instructed to produce as many different concepts as possible. In no investigation of problem-solving can we afford to ignore the role played by the instructions given to the subjects (Wertheimer, 1945, seems to be guilty in this respect).

Limiting the present discussion to practical construction problems of the Maier type, it may be said that the subject has a goal which he must attain by using a definite number of objects in definite ways and that these objects serve definite 'functions' in the attainment of the goal. If problem-solving is thought of in terms of these 'functions', there arises at least one important question. When a problem is not solved, is it simply the result of one or more of the necessary 'functions' being unavailable to the organism, or may it also be due to some difficulty in combining these 'functions'? Or, otherwise expressed, if a subject has available the necessary 'functions' for the solution of the problem, will he actually solve it?

An answer to the above question demands that the availability of 'functions' be determined independently of the subject's behaviour in the problem-solving situation. This determination may be made by presenting the subject with the objects which are later to be used in solving the problem and asking him to list all the 'functions' which he thinks each object could have. In this way, there may be obtained in advance subjects who have available the necessary 'functions' for the problem's solution. Which particular 'functions' are or are not necessary depends, of course, on the exact nature of the problem later to be set. Furthermore, there is no guarantee that a subject who fails to list a particular 'function' beforehand may not get the 'function' when presented with the problem situation. The main interest of the following experiment will therefore centre round the group of subjects who give, in advance, all the necessary 'functions'.

Before going on to describe the actual experiment, it should perhaps be indicated what the writer regards as the three main requirements of any problem situation in which the role of the availability of 'functions' is to be examined. First, there should be several objects and thus several 'functions' to combine. With too few 'functions' it might be argued that the situation did not constitute a fair test of the role played by the combination aspect of the solution. It is noteworthy that the problems usually investigated in this field rarely involve more than just a few 'functions'. Secondly, the problem must not be presented in such a way that the subject doubts, for some trivial reason, the appropriateness of an object to any necessary 'function'. If a rod, for example, has to be used for horizontal support, the objects to be supported must not be so heavy that the subject comes to doubt whether the rod is strong enough to carry their weight. Thirdly, if a time limit is imposed, it is important that the subject be able to discover fairly rapidly that erroneous modes of attack are, in fact, wrong. A wrong, but plausible seeming, approach may so engage the subject that he never attempts the correct solution of which he may well be capable. More than one correct solution might also complicate matters, for, having found one, especially if it has been found with difficulty, the subject will tend to rest content with it and not attempt the other.

II. EXPERIMENTAL

(i) *Subjects*

The subjects were 205 male college-student volunteers, aged 17–20 years. Of these, 120 took part in a pilot investigation designed to standardize the problem and the instructions. The experiment proper involved 85 subjects of whom 57 were allocated randomly to an Experimental Group and 28 to a Control group.

(ii) *The problem*

The problem was a modification of Maier's (1933) candle problem. A candle, about 6·5 in. high, had to be blown out from a distance of 70 in., the candle being placed on one table and the subject standing behind another. Available to the subject were: six hollow glass tubes, 9·5 in. long and $\frac{3}{4}$ in. in outer diameter; a solid duralumin rod, 50 in. long and $\frac{5}{16}$ in. in diameter; a 1 oz. lump of putty. The particulars of the materials used are of importance since the pilot study revealed that they influenced both the difficulty of the problem and the frequency of listing the various 'functions'. Each subject was given the following instructions. 'Here is a problem for you. I want to see if you can put out this candle from behind this table. I want a safe good method. If you do not feel like doing it, you may explain to me how you would do it. You may use this material [the experimenter pointing to the material on the table] or not'. If the subject suggested an incorrect approach to the problem he was not informed that it was wrong but asked to perform what he had in mind. If he tried to lengthen the rod either by stretching across the table or by fixing a glass tube on to the end nearest him, he was told that one end of the rod had to be in line with the edge of the table nearest him. It might have been possible for the subject to solve the problem by lengthening the rod, through puttying glass tubes on to its end, and then tapping out the flame. To prevent this, the cohesive power of the putty was lowered by moistening it slightly and the glass tubes were brought in wet after newly having been washed. The latter device also served to reassure the subject that the glass tubes were clean enough to be touched by his lips.

Under the above conditions, the only possible solution was to put the glass tubes end to end round the rod and, by means of the putty, make the joints between them air-tight. By kneeling down the subject could then lift the blow-pipe structure (preferably while leaning his elbows against the table), direct it at the flame and, by blowing through it, extinguish the candle.

The pilot study showed that not all the subjects who saw the solution were able to manipulate the practical details of it. The following criterion for the correct solution was, therefore, adopted: the subject must place at least five tubes on the rod, seal up at least one joint completely, and either blow or clearly express an intention of so doing. Merely placing a lump of putty across the joint of two tubes was not considered sufficient since subjects often did this preparatory to sticking the tubes together in line rather than to sealing them into one long airtight tube. A subject might attempt to bring the end of the construction that was nearest to the candle to the same level as the flame by pressing down on the table with the end nearest him. In this case, he was told that the tubes would break.

(iii) *Procedure*

Each of the 85 subjects was presented with the candle problem as described above. The time limit was 30 min. and if a subject wanted to give up before then he was encouraged to continue until the time was up. The Experimental Group differed from the Control in that it was given two previous sessions in which the availability of the necessary 'functions' was tested for.

The first session given a subject in the Experimental Group lasted for 15 to 30 min. and was designed to make clear to him what was required in the availability test proper. The necessity of including this session was made clear by the pilot study. The subject was told that he would be shown a number of objects and that he was to list all the 'functions' which he could think of in connexion with each. The objects were not to be thought of as changeable, i.e. as being alterable by, say, breaking or melting. When he had a 'function' in mind, he was to state, first the physical characteristics of the object which allowed it to serve this 'function' and, secondly, the 'function' in a general form giving three concrete examples. He was told to keep the listing of each 'function' separate from the listing of every other function. The practice object was an iron bolt. During this session, the subject was free to question the experimenter on any points of difficulty.

The availability test proper lasted for 13 to 30 min. The objects necessary for the solution of the candle problem were presented in the following order: the glass tube, the lump of putty, the rod. The three objects were shown one at a time and, for each, the subject listed its 'functions' until he could think of no more. The 'functions' were stated verbally and noted down by the experimenter. Those 'functions' which would later be necessary for the solution of the candle problem were three in number: (1) the glass tube had to conduct an air stream when blown through, (2) the putty had to make crevices, apertures, joints, etc., either air- or water-tight, and (3) the rod must support objects hung from it when held in a horizontal position. A 'function' was judged as given when it was stated in a general form accompanied by one appropriate concrete example. The subject's statements relevant to the three necessary 'functions' were given to two psychologists who scored them independently of the experimenter and of each other, Subjects in the Experimental Group were given the candle problem immediately after completing the availability test.

(iv) *Results*

As stated in the Introduction, the experiment's chief interest was the performance of these subjects who, in the availability test, gave the three 'functions' of 'blow', 'make tight', and 'hang'. However, after testing 35 subjects in the Experimental Group, it was found that only one gave all three 'functions'. Yet every one of seven subjects who gave only the second and third 'functions' solved the problem. It was decided accordingly to continue the experiment by regarding 'make tight' and 'hang' as the necessary 'functions' for otherwise a very large number of subjects would have had to be run.

Of the 56 subjects in the Experimental Group[1], 13 (group 1) gave the functions of 'make tight' and 'hang'. All 13 solved the problem. There were 43 (group 2) subjects who failed to give both 'functions', and only 25 of these solved the problem. If the difference between those two groups, with regard to the number of solutions, is tested by Fisher's exact method, a P value of 0·003 is obtained on a one-tailed test of significance. Of the 18 subjects in group 2 who did not solve the problem, 11 did not use the rod for hanging and 7 placed the glass tubes on the rod and attempted to blow out the candle without sealing the joints.

The two psychologists who scored the statements disagreed on one statement relevant to the function 'hang' and on four relevant to 'make tight'. Only subjects on whose statements there was agreement were included in group 1.

The performances of the same 56 subjects were also assessed on the basis of a slightly less rigorous statement of the two necessary 'functions'. They were credited with the availability of both 'functions' if their comments merely implied that: (1) the putty might make tight in some way (e.g. tight against water, air, light, or insects; covering holes to obtain a smooth surface was not acceptable), and (2) the rod might act as a horizontal support. In terms of those criteria, 27 subjects (group 1) had both 'functions' available and 23 (group 2) had not. In group 1, the problem was solved by 23 subjects and in group 2 by only 11. If the difference between those two groups is tested by a χ^2 corrected for continuity, a P value of 0·006 is obtained on a one-tailed test of significance. (It will be noted that six subjects have been excluded from these results. This is because the two scorers could reach no agreement as to whether or not those subjects had given both necessary 'functions' in the availability test.)

---

[1] One subject was left out of the material as, after 10 min. he managed to blow out the candle without making the joints tight with putty, and was unwilling to continue, arguing that he had solved the problem. This subject's solution is a little doubtful. He placed five tubes on the rod and applied putty almost all round one joint, but not on the lower part of the joint. When asked after the session why he had not made the joints tight, he contended that that was what he had been trying to do, but there was no reason to seal on the lower part of the joint as no air would escape there. In contrast to two other subjects who came so far towards the solution that they placed a lump of putty over a joint, this subject maintained that he had intended to make the joints tight and not just to try to stick two tubes together by the putty. Putty was never used for glueing purposes, he said. Both his overt and verbal behaviour seem to be different from that of the other two subjects.

Of the 28 subjects in the Control Group, 16 solved the candle problem. When this number of solutions is compared with that of the Experimental Group by means of χ^2 corrected for continuity on a two-tailed test of significance, the P value is 0·47. Thus there was no difference in the ability of the two groups to solve the problem. Nor was there any difference between the groups with regard to the time required for the solution: the successful members of the Experimental Group required an average of 10 min. 30 sec. for solution, and those of the Control Group an average of 10 min. 33 sec. Observation suggested, however, that there may have been qualitative differences in the problem-solving behaviour of the two groups. If the correct solution is regarded as a definite number of 'functions' performed in a definite sequence (i.e. blowing with the mouth, blowing through one tube, blowing through two tubes, putting the tubes on the rod and blowing, and then sealing and blowing), the Experimental Group seemed to perform in a less smooth sequence than the Control and with more trial-and-error. But a detailed discussion of those qualitative differences cannot be given here.

III. DISCUSSION

The above results seem to indicate (1) that subjects able to give the necessary 'functions' in the availability test are later able to solve the set problem, and (2) that the availability test, as such, may have little or no effect on the subsequent problem-solving behaviour. They suggest that the difficulty of any problem rests in the unavailability of one or more of its necessary 'functions' rather than in any difficulty of combining these 'functions' together. Further, it would not seem necessary that all the necessary 'functions' be combined at the one time; it may be that most problems can be broken down into sub-problems involving a smaller number of 'functions' than does the original.

The critical theoretical issue of the present paper concerns the essential difference between a subject who solves a problem and one who does not. Are they to be thought of as (1) differing in the availability of necessary 'functions', or (2) differing in the possession of an 'ability to combine functions'? In so far as the above experiment succeeded in establishing objective criteria for the necessary 'functions', for the availability of these 'functions', and for the solution or non-solution of the problem, the first alternative seems the more plausible. After all, there is, as far as the writer knows, no clear evidence of the existence of an 'ability to combine functions'. But if there were such an ability, it might be possible to explain the above results in terms of it. Thus it could have happened that the availability test selected, for group 1, subjects of superior 'verbal ability', and this ability might correlate highly with that required in 'combining functions'. In that case, group 1 would comprise a selected group of subjects of superior 'ability to combine functions'. On the other hand, it would be equally admissible to argue instead that the 'verbal ability' correlates highly with ability to make 'functions' available. Another way in which an 'ability to combine functions' might vitiate the present interpretation is as follows: If the students participating in the experiment all had a high 'combining' ability, then the only feature which would differentiate among them would be the availability of the necessary 'functions'. This argument, however, appears to return us to the original interpretation. It might, of course, be that 'verbal ability' does not correlate too highly with either 'ability to produce functions' or with 'ability to solve problems'. This is perhaps suggested by the fact that when the criterion of availability was made less rigorous (i.e. when the subject need only imply a reference to the necessary 'functions') the results seemed little altered.

It should be made clear that the writer does not deny an 'ability to combine functions'. He merely asserts that such an ability has not yet been established, as some psychologists appear to imagine, and wishes to draw attention to it as a problem for future research.

So far we have avoided any discussion of what may be involved in the availability test. Did it make the combination of the 'functions' easier, or more difficult? As long as we know so little about what processes take place when someone produces a 'function' in the availability test, it is difficult to answer this question or even to state what is meant by 'combination of functions'. Comparison of the Experimental and Control Groups does not reveal that the availability test gave the Experimental Group any advantage over the Control with regard to solving the candle problem. It is, in fact, difficult to see how the procedure in the availability test could make any one combination of 'functions' easier than any of the other many combinations possible with regard to the numerous 'functions' of the three objects. Indeed, it is not impossible that, by making available a larger number of 'functions', the availability test might actually make the solution of the problem more difficult. There is, too, the possibility that a 'function' might be available during the availability test and yet, for some reason, be unavailable in the actual problem situation.

In concluding we may comment on the advantages to be gained if our interpretation in terms of availability of 'functions' is valid. In the first place, it would seem to imply that the number of 'functions' that have to be combined does not necessarily determine the difficulty of the problem. As long as the necessary 'functions' are available, it probably does not matter greatly whether there are few or many. This implication might help to increase our understanding of what is meant by the term 'problem' and also help to integrate psychological material which, at present, tends to be dealt with under such separate headings as 'learning', 'problem solving', 'memory', 'perception', 'motivation', and 'intelligence'.

Whether or not the present approach is fruitful is a matter of the usefulness of the questions which it raises. For example, are 'functions' combined by being added together in successive fashion? Does the solution develop as a systematic trying out of all available 'functions' until the correct pattern is obtained? Would it be correct to predict that, in any problem situation, the ability to produce 'functions' is highly correlated with the ability to solve problems?

IV. SUMMARY AND CONCLUSION

When a problem involves a goal to be reached by the manipulation of concrete objects, these objects may be said to serve certain 'functions' in the attainment of the solution. Eighty-five college students were set a modified version of Maier's candle problem. For 57 of these subjects the problem was preceded by an availability test. In this test nothing was mentioned of the problem itself, but the objects later to be used in it were presented one by one and the subject instructed to list all the possible 'functions' which the objects

might serve. By this technique, a group of 13 subjects was obtained who, in advance of the problem situation, had given evidence that the necessary 'functions' were available. All 13 subjects later solved the problem. Of the subjects who did not indicate that the necessary 'functions' were available, only 58% solved the problem. Stated in the most general terms, these results are interpreted to mean that a subject will solve any problem if he has available the necessary 'functions' and that it is not the difficulty of combining these 'functions' which renders a problem insoluble for a particular subject. The implications of this interpretation are briefly discussed.

The writer wants to express his indebtedness to his sponsor L. L. Thurstone, to E. H. Porter Jr., Austin H. Riesen, Garth J. Thomas, Donald W. Fiske and Lyle V. Jones, and to Dr I. M. L. Hunter (University of Edinburgh) for invaluable help with the manuscript.

References

Harlow, H. F. (1951). Thinking. Pp. 452–505 in *Theoretical Foundations of Psychology*. (Editor, Harry Helson.) New York: Nostrand.

Humphrey, G. (1951). Thinking. London: Methuen.

Maier, N. R. F. (1930). Reasoning in humans: I. On direction. *J. Comp. Psychol.* x, 115–43.

Maier, N. R. F. (1933). An aspect of human reasoning. *Brit. J. Psychol.* xxiv, 144–55.

Maier, N. R. F. (1937). Reasoning in rats and human beings. *Psychol. Rev.* xliv, 365–78.

Maier, N. R. F. (1945). Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. Exp. Psychol.* xxxv, 349–60.

Wertheimer, M. (1945). *Productive Thinking*. New York: Harper.

Weaver, H. E. & Madden, E. H. (1949). Direction in problem solving. *J. Psychol.* xxvii, 331–45.

(*Manuscript received* 28 *May* 1954)

A STUDY OF THE PROBLEM OF INTELLECTUAL DETERIORATION IN ADOLESCENTS AND YOUNG ADULTS

By K. LOVELL

Borough Road College, University of London Institute of Education

I. *Introduction* (pp. 199–200). II. *Outline of previous research* (pp. 200–202). III. *Plan of the present investigation* (pp. 202–205). IV. *Discussion and results* (pp. 206–208). V. *Summary* (p. 209). *Appendix* (p. 209). *References* (p. 210).

I. Introduction

Some decrease in intellectual efficiency is one of the commonest manifestations of mental disorder, cerebral lesion and age, while the term deterioration is generally taken to mean a progressive change in the direction of less and less efficiency. Moreover, it is usually implied in the literature that deterioration involves an irreversible change, although this is seldom stated explicitly. In this study it must be understood that the term is used in a more neutral sense, since we cannot be sure if the levels of performance found in the different groups taking part in this investigation could be improved upon given an environment more intellectually stimulating. Furthermore, it is impossible to say if the levels of performance found indicate a decline compared with those that would have been found earlier, or whether some abilities in certain of our groups never came to fruition.

The problem of intellectual deterioration among normal adolescents and young adults is one which has been studied but little. Psychologists and psychiatrists have certainly made great efforts to measure intellectual deterioration in populations suffering from certain organic and psychopathological conditions, but among younger, non-clinical cases, the problem has been largely ignored, while the findings among clinical cases to date have been somewhat confused and contradictory. This lack of clear findings is thought to be due in part to the failure of research workers to pay proper attention to the design of their experiments and to the basic principles of mental testing. Again, the dearth of experimentation among normal young people conducted along lines similar to those used in the case of clinical subjects is, presumably, due to the fact that it was considered that intellectual deterioration did not occur among the former. Now, however, it is clear that general intelligence begins to decline in a normal (i.e. non-clinical) population at a comparatively early age, and the rate of decline seems dependent both upon the original level of intelligence and upon occupation, while verbal educational abilities are certainly affected by teaching and environmental opportunities generally. Differences in level of performance in respect of some other abilities might therefore conceivably be found among groups with varying backgrounds. Furthermore, it is possible that such differences are the same as some of those said to be indicative of changes in efficiency associated with the onset of certain clinical conditions.

Vernon & Parry (1949) summarized the position as it was up to about 1947 in relation to the effects of age on intelligence test scores. Norris (1940), in an extensive investigation, found that educational attainments tend to lapse rapidly after leaving school except in so far as they are used in daily life. Again, Vernon & Parry (1949) suggest that verbal educational abilities (other than reading ability which shows an improvement) seem either to remain stationary or to decline over the period 14 to 18 years of age unless they are stimulated by secondary or further schooling. As a result of comparing scores obtained by Glasgow school leavers and Glasgow naval recruits they submit that spatial mechanical abilities may rise over this period whether or not further education has been received.

Over and above these well-established *g*, *v*, *k* abilities, which are of such proved worth in education and in vocational guidance, there are many notions which have, in recent years, been discussed and developed in the literature concerning 'concept formation', 'rigidity' and the like. It is true that these terms have mostly arisen out of research with clinical cases of one type or another and the results are, to say the least, somewhat confused. Nevertheless, if factors of 'concept formation' or 'rigidity of thinking' exist, over and above *g*, *v*, and *k*, these should be investigated, since it is possible that deficits found among mental patients may also occur in young, non-clinical cases. Typical of the viewpoint that concept formation is inadequately measured by current tests of intelligence is that expressed by Hearnshaw (1951). After condemning the view of Burt and Vernon that the map of the intellect is more or less complete, he argues that he suspects that into the process of concept formation there may enter a type of abstraction that enters only incidentally into conventional tests of intelligence. 'These tests', writes Hearnshaw, 'involve for the most part simple relation and correlate eduction, not the ordering of manifold and heterogeneous experience through the grasp of abstract principles of order. Thus Matrices presents a determinate system of relations one term of which is missing. This is quite different from a determinable principle of order which can be variously exemplified. The Matrix is relationally concrete in comparison with the concept.

II. OUTLINE OF PREVIOUS RESEARCH

The literature surrounding this study is extensive; mention can be made of only a few of the more important references. Goldstein & Scheerer (1945), Hanfmann & Kasanin (1937), Berg (1948), and others, have devised a number of concept formation or sorting tests. In a typical test the subject is presented with a series of objects which can be classified in various ways and he is told to sort them according to a given principle or pick out others similar to a given object. When the subject has mastered one principle of classification he is given a new and more complex problem. It is claimed that brain-injured patients and some schizophrenics can cope with these at the concrete level, but cannot realize and formulate abstract principles of classification, nor can they switch from one principle of classification to another.

A further concept to be considered is that of perseveration, the tendency to inertia in mental processes, as contrasted with the ability to switch quickly from one process to another. Walker, Staines & Kenna (1943), and Cattell (1946) now state that the earlier tests so often used to measure perseveration (e.g. the alternation type of test and miscellaneous sensory tests) are useless; but that 'creative effort' tests, where the subject has to break down some well-established habit, are more consistent and measure a quality which has been termed 'disposition rigidity'. Cattell claims that a person with low disposition rigidity is a person of strong ego development and capable of modifying his habits. Little, however, seems to be known of the validity of tests of disposition rigidity.

Luchins (1942, 1947) maintains that inefficient methods of education may produce rigid rather than flexible minds and has devised a number of paper-and-pencil tests.

Subjects are given a series of problems which can be solved by the application of one principle. Half way through the principle alters, but the original 'set' may delay or prevent them entirely from realizing this. The interference in such tests is established temporarily during the course of the test, whereas in disposition rigidity tests the interference is with a fully automatized habit. Yet another concept of rigidity has been put forward by Lewin (1935) and Kounin (1941). These and other Gestalt-minded psychologists regard rigidity as one of the main characteristics differentiating the personalities of the feeble minded from those of normal children. When given a choice of drawing tasks, feeble-minded children tend to persist longer at a single task, whereas normal children prefer a change.

Again, there is a mass of literature dealing with tests of the type included in the Wechsler–Bellevue Battery (1944), the Babcock–Levy Battery (1940) and the Shipley–Hartford Test (1940). The performance of psychotics, especially those of the organic type, is poorer than expected on some of the tests; performance also declines in aphasic conditions, in cases of brain lesion and to some extent with age. The decline is least marked in tests of what Cattell calls 'crystallized' intelligence, such as vocabulary. Tests of 'fluid' intelligence, which are most affected, include speed tests and tests of abstraction or conceptualization.

Few attempts have been made, until the present inquiry, to determine whether tests such as those described by Goldstein & Scheerer, tests of disposition rigidity, and tests of the Wechsler–Bellevue Battery, in fact measure anything over and above the well-established *g*, *v*, *k* abilities. Absence of such work may be due to the following considerations:

(*a*) Most of the relevant work in this field has been done in America where there is less agreement on a small number of ability factors.

(*b*) Clinical workers, and those generally who have done most of the work in this field, seem often to be hostile to a factorial approach, or else ignorant of or indifferent to the value of such an approach.

(*c*) Since the available tests have been individual ones there would have been an enormous amount of labour expended in the giving of a battery of say concept formation tests to a group of a size adequate for a factorial approach. The one serious attempt made to use factor analysis was that of Halstead (1947) who applied nineteen tests to fifty patients. Neglecting the analysis which led to an oblique factor structure and concentrating on the analysis performed by Holzinger which gave a general and orthogonal group factor solution, it seems clear that the general factor is general intelligence or *g*, while Halstead's A factor is possibly the ability to group to a criterion for it has its highest loading 0·6 on a category test. This work has been criticized by Vernon (1950).

Crown (1951), in a review of the literature dealing with psychological changes following prefrontal leucotomy, comments on the lack of consistency between investigations on the issue of whether or not abstraction is affected following the operation. This lack of agreement is not surprising, he maintains, considering the variety of tests which have been used to measure abstraction. He divides these tests into two main types:

(1) Those which are primarily intelligence tests—Kohs Blocks, vocabulary tests, Shipley, proverbs, fables and abstract words.

(2) Tests which appear to conform to Halstead's definition of abstraction as the basic capacity to group to a criterion. Tests of the latter type involve the comprehension of essential similarities and differences—as in, say, the Goldstein–Scheerer Colour Form

Sorting Test. 'It is suggested', writes Crown, 'that, in factorial terms, sorting tests, if included in a test battery with other cognitive tests, would cluster together, these tests having, in addition, moderately high correlations with intelligence. Whether the abstraction group factor would be different from group factors already widely recognized remains a problem for further research.' In the opinion of the writer Crown's suggestion comes remarkably near the truth.

III. PLAN OF THE PRESENT INVESTIGATION

The general plan of attack on the problem of intellectual deterioration in adolescents and young adults adopted by the writer was as follows. A review was made of a representative selection of the literature dealing with the ability to think abstractly, to conceptualize and to switch from one conceptual set to another in psychotics, organics, neurosurgical cases and normals. Linked with this was a study of the literature dealing with motor and cognitive rigidity. From the overall field a selection of tests was made which appeared to be best substantiated by past research. To these tests was added a test of non-verbal intelligence and one or two of the more usual tests used in educational research, i.e. reading and arithmetic. The test battery was applied to various groups of young people, all non-clinical subjects, between the ages of 14 and about 20 years, and the results submitted to factor analysis to discover whether in addition to g, $v{:}ed$ and $k{:}m$ abilities any further group factors could be found such as:

(*a*) ability to reason at the abstract or concrete level.

(*b*) ability to switch from one conceptual set to another.

(*c*) ability to switch from one set of habits to another in the motor or cognitive fields.

Using point biserial correlations in connexion with group differences it might be possible to validate any additional factors found, against an external criterion.

The experimental population consisted of an older and a younger group. The testees in the latter category were still at school and were in turn divided into two groups, one of which came from a co-educational Secondary Modern School where the atmosphere was considered to be stimulating. This group consisted of boys and girls in their last year at school. The other group, which consisted of boys only, also in their final year, came from a Secondary Modern School where the atmosphere was reckoned to be non-stimulating. It is admitted at once that the criterion of stimulation was based on subjective judgements, but the judgements were made without dissent, by Training College tutors who had had good experience in visiting different types of school. To be specific we may say the school which provided a stimulating atmosphere was characterized by:

(*a*) a good standard of teaching maintained by the staff employing in the main a formal approach to teaching;

(*b*) A good attitude to work on the part of the pupils whether one considered the more academic subjects or, say, music or craft;

(*c*) An excellent spirit existing between the children and the staff.

It is a truism to say that the school often reflects the neighbourhood, but this point must be stressed. In areas where the children come from good stock and helpful homes there also the more able teachers tend to congregate and stay. As a result, the life and work of the school, already being built on potentially sound material, are further enriched. Thus it is quite impossible to disentangle the school and environmental influences, and we must leave our discussion of the degree of stimulation afforded by the background of our school groups by saying that the stimulated group not only came from a better school, but most probably from an intellectually healthier environment and more helpful homes, and that the children within the group were rather more stable in their behaviour. The average age of both of the school groups was close to 14½ years, and it must be emphasized that the economic conditions within the groups was about the same as far as could be judged.

In the other half of our experimental population there was, once again, a stimulated and non-stimulated group. The former group was made up of young men and women who were working in factories, offices or at trades, and who came to a Day Release Centre for one day per week. The members of this group had the following characteristics:

(*a*) They were all employed by progressive firms or organizations that made attendance at a Day Release Centre a condition of employment;

(*b*) They attended a Day Release Centre where the standard of teaching was assessed as good, where the attitude of the students to work was sound and where staff-student relationships were excellent.

(*c*) They had been exposed to the influence of work and of the Day Release Centre for as long as possible before testing.

(*d*) They were not the most able people attending the Centre (our purpose was to get a group comparable in ability to our school groups).

The very fact that this section of our population was holding down posts with progressive firms or organizations would indicate that it was composed of people who had some liveliness of mind and some stability of behaviour, for the bad work record of the delinquent and maladjusted is well known. The average age of this group was 17½ years.

The problem of getting an older non-stimulated group is a formidable one, principally because normally one has no hold over such a group which would make it submit to a testing programme that lasts several hours. Several possibilities were explored. Eventually through the kindness of the Prison Commissioners[1] the difficulty was overcome, permission being given to test a group of young prisoners at a rather special type of prison in the south of England that has been set aside for such offenders. To be eligible to serve a sentence there the prisoner must be under 21 years of age when sentenced, and the sentence must be for a period not greater than four years. Our request was for a batch of prisoners who had been admitted to the prison within a few weeks of the opening of the testing programme. Thereby it was hoped that it would be possible to examine these young men literally as they 'came off the streets', and before prison work and routine had had any appreciable effect on them (some, of course, had been to an Approved School or Borstal Institution or both). A further request was made that the outstandingly able, the very dull and all illiterate prisoners be excluded from the test group, as it was hoped that it would be possible to make the group comparable with the other groups as far as ability was concerned. Our total population thus consisted of four groups of approximately fifty subjects in each group, making some 200 cases for examination in all.

The conditions governing the choice of tests were as follows:

(*a*) The tests had to be representative of the many different types that have been used in the field of mental deterioration, and at the same time they had to be among those that appeared to be most promising.

---

[1] Any opinions expressed do not necessarily represent official views.

(*b*) It had to be possible to put all the tests into group form. The Vigotsky Blocks Test was ruled out on this account.

(*c*) The overall testing time, including time devoted to reading out the instructions and ensuring that every one of the testees understood what had to be done, could not be more than about eight hours.

(*d*) Bearing this maximum in mind it was essential to devise group versions of the chosen tests which would yield reasonable reliabilities, and which might therefore obtain reasonable communalities in the factor analysis.

When appropriate tests had been selected or prepared, all, apart from well-known tests, were tried out on the older age groups of a Secondary Modern School (not one of the two schools from which the experimental groups came). In some of the cases the time allowances, level of difficulty of the items, etc., seemed to be about right on the first trial, but in other cases considerable alterations had to be made and the revisions tried out again on different but comparable children. Although it had been hoped that the trial groups of children were intellectually comparable with the experimental groups, it is the writer's opinion that the latter were more homogeneous with respect to ability than the former. This means that our correlations are somewhat lower than those to be expected from a more representative sample. The tests finally included are now listed with a few relevant comments:

Test 1. N.I.I.P. Group Tests 70/2 and 70/3.

Test 2. Watts–Vernon Reading Test.

Test 3. Arithmetic Test. This consisted of fifty problems similar to the problem section of the Moray House type of Arithmetic Test and thus similar to the Wechsler–Bellevue Arithmetic Test, but much longer.

Test 4. A shortened version of the Wechsler Vocabulary Test.

Test 5. A group version devised by the writer of the Wisconsin Sorting Test. This test was scored in two ways—for the principle of classification ('labels'), and for sorting the pieces correctly ('pieces').

Test 6. A group version devised by the writer of a sorting test suggested by Vinacke (1951). This test was scored in two ways as in the case of Test No. 5.

Test 7. A group test prepared by the writer based on the Trist–Hargreaves test (Semeonoff, 1955). This also was scored for the principle of classification as stated, and for the pieces correctly classified.

Test 8. A test of Concept Formation of the Heidbreder (1949) type.

Test 9. A non-verbal Classification Test. The test was scored for the underlining of the correct figures and for stating the principle of classification.

Test 10. Proverbs.

Test 11. Test of Memory for Stories.

Test 12. Babcock Test of Memory for Sentences.

Test 13. Test of Memory for Designs including fourteen of the designs used in the Bender Visual Gestalt and Ellis Tests.

Test 14. Creative Effort Test. The writing of ZEK forward and then in reverse as ƧƎꓘ.

Test 15. Creative Effort Test. The writing of 239 forward and then in reverse.

Test 16. Creative Effort Test VAT. The letters were written first in the normal way and then upside down.

Test 17. A Creative Effort Test in which there were 46 very simple addition and subtraction exercises but whenever there was a plus sign the testee had to subtract and vice versa.

Test 18. Similar to Test 17 but consisting of interchanged multiplication and division exercises.

Test 19. Creative Effort Test—Seasons. Here the subject had to associate a month of the year with a season, e.g. August with summer. But sometimes the process had to be reversed, e.g. April had to be associated with autumn.

Test 20. This was the first of three tests which in themselves are entirely original although of the type used by Luchins. It involved the simplification of very easy fractions and was in two parts the first of which measured basic speed of working.

Test 21. Hidden Words. The test was in two parts, the first of which measured the basic speed of working.

Test 22. Beginnings and Endings of words. Again the test was in two parts, the first measuring the basic speed of working. In each of tests 20, 21 and 22 there was an attempt to develop a mental set in Part I of the test and to measure the extent to which this set interfered with the working of Part 2.

Test 23. Three different operations were first performed separately and afterwards similar operations had to be performed in quick rotation. It was suggested by the Army Clerical Test (Vernon & Parry, 1949).

Test 24. Gottschaldt Figures: 42 items from Thurstone's revision.

Test 25. An extension of the Shipley Test made by Vernon.

Test 26. A group version of the Kohs Blocks Test devised by Hey (1951).

In the case of Tests 14–16 and 20–23, basic speed scores were graphed against the rigidity scores (i.e. the scores obtained in the second part of the first six tests listed above, and in the fourth part of Test 23) and the approximate regression line drawn in. The rigidity of a particular case was then read off by subtracting from the obtained rigidity score, the rigidity score indicated by the regression line corresponding to the basic speed score. Rigidity was thus expressed in terms of a positive, zero or negative quantity and was made independent of the basic speed score. This appears to obviate criticism often levelled against earlier workers that the estimate of rigidity was dependent on basic speed; it also appears to solve a difficulty which Luchins and his followers seem to ignore.

All raw scores were first recast so as to make the distributions normal and product-moment correlations calculated after allowing for differences in subgroup means. Thurstone's Centroid Method of Factor Analysis, with subsequent rotation to a general and orthogonal group factors solution, was the method of factor analysis considered to be most appropriate to this experiment, while every effort was made to establish the well authenticated g, v and k abilities first. Finally, Point Biserial correlations were calculated in the case of each test for Stimulated versus Depressed or Non-Stimulated groups and for Youth versus Age groups. These correlations, however, were not included in the matrix; instead, correlations of 'Stimulation' and 'Youth' with the found factors were later calculated.

IV. Results and discussion

An examination of the means of the raw scores for the various subgroups revealed at once that the Stimulated groups did relatively much better on some tests than on others, and it was further established that this was not due to sex (it will be recalled that there were females in the Stimulated but not in the Depressed groups). For example, the Stimulated groups had relatively better scores on the Sorting Tests than on the Reading Comprehension Test. Accordingly, all group differences were reduced to standard scores. On adding together the mean standard scores of the two Stimulated groups and subtracting from this total the sum of the standard scores obtained by the Depressed groups (i.e. for each test in turn) it was seen at once that there was most deficit on the six concept formation variables and least on tests of reading, vocabulary and Group Test 70/2. A few results are given below; they strongly suggest at least that a form of deterioration may be present in the concept formation type of test not normally detectable in *g* and *v*:*ed* tests:

Test	Difference between stimulated and depressed groups	Test	Difference between stimulated and depressed groups
Three Activities	0·80	Kohs Blocks	−0·20
Wisconsin Pieces	0·69	Shipley Abstraction	−0·43
Vinacke Labels	0·62	Wechsler Vocabulary	−0·59
Wisconsin Labels	0·58	Reading Test	−1·01
Test of Fractions	0·62	N.I.I.P. 70/2	−1·06

The results of the factor analysis are set out in the Appendix with the factor loadings correct to two decimal places and with loadings less than 0·14 omitted. Six of the seven rotated factors obtained are, in the opinion of the writer, as clear-cut and unambiguous as can be expected from a factorial study involving tests in a new field. The first factor can be described as *g*, or general cognitive capacity, and without doubt indicates the presence of an ability which is responsible for most of the identifiable variance in non-verbal intelligence tests such as Matrices and Dominoes. The second factor is indicative of verbal ability; an ability of proven worth in educational and occupational prediction. The highest saturations of this factor occur on the Watts Vernon Reading Test, the Wechsler Vocabulary Test and on Proverbs. The third factor is clearly the spatial factor *k*; it correlates most highly with Memory for Designs, Gottschaldt Figures and the Kohs Blocks Tests. With the fourth factor this research appears to be breaking new ground. That is to say, factor four seems to be indicative of an ability long suspected by clinicians to be present in non-clinical cases but absent, in varying degrees, in certain clinical cases (i.e. brain lesion cases, organic psychotics and some functional psychotics). The factor has its highest loadings on the Sorting Tests, the test of non-verbal categorization or classification, and the three Luchins type tests. Note that the actual maximum loadings were obtained by tests 9 B and 6 L respectively, where in both instances, the subject has to say precisely how he has made the groupings. It is suggested that this factor should be labelled *c* for categorization (cf. Goldstein's notion of the categorical attitude) and that the core of the ability indicated by this factor seems to be the ability to group objects or ideas according to a criterion and to switch from one criterion to another. There is therefore a definite element of flexibility linked with this ability; indeed it is likely that the *c* factor is rather more complex than our analysis was able to bring out.

The fifth factor has been labelled non-perseveration (motor), since it has its highest loadings on motor perseveration tests of the creative effort type. It is suggested that this factor may also be indicative of Cattell's disposition rigidity. Factor six is indicative of what Vernon (1950) has described as a numerical educational ability (n:*ed*). Summarizing, it may be said that the analysis has yielded the well-established g, v, k and n:*ed* abilities, a non-perseveration (motor) factor, and most important, it has yielded evidence of a factor indicative of the ability to categorize (independent of these other five abilities), long suspected but not previously demonstrated to the satisfaction of British psychologists.

A major criticism often levelled against the work of factorialists is that the found factors often seem to be little more than statistical artefacts devoid of any 'real' existence. In an attempt to overcome this criticism an effort was made to validate our analysis by correlating the variables of Stimulation and Youth with the found factors. The variable Youth is easy to interpret, it represents the difference in scores between the younger groups (i.e. the two school groups) and the older groups (Prisoners and Day Release students). The variable Stimulation, on the other hand, cannot be defined with such exactitude. There is no evidence that it is covered completely by the quality of the school teaching that the two groups (i.e. the stimulated and depressed groups) received. This may have contributed something but it is thought that the general background, attitude to life and work and the general adaptability of the two groups differed considerably. Thus the Stimulation variable will depend upon inherited traits, structure of the central nervous system (in the sense of being able to withstand a certain amount of stress and make adequate responses to real life situations), early upbringing and environment, cultural background, quality of teaching and school atmosphere generally.

The correlations between the two criteria and the six factors are now given with coefficients less than 0·14 again omitted:

	g	v	k	c	non-p	n:*ed*
Stimulation	—	−0·18	−0·24	0·50	—	—
Youth	0·17	−0·25	—	—	—	—

The following points emerge:

(*a*) There seems to be no relationship between g and the Stimulation variable. In other words, the stimulated and non-stimulated groups were matched for general intelligence.

(*b*) The non-stimulated group seems to have slightly higher verbal ability.

(*c*) The non-stimulated group did better on tests involving a large amount of k ability. This is precisely what one would expect for it has often been pointed out that delinquents do relatively better on k tests than do normals; also the stimulated group contained females, whereas the non-stimulated group did not.

(*d*) The substantial correlation between the Stimulation variable and the c factor provides strong evidence that this factor is not merely a statistical artefact. When it is remembered that our criterion cannot be a 'pure' one, the size of the correlation between the factor and the variable appears very encouraging. It suggests that our non-stimulated or depressed group is behaving very much like Goldstein's brain-injured patients.

(*e*) Verbal ability tends to rise between 14 and 20 years of age largely because vocabulary and reading ability increase. This confirms the data presented by the Ministry of Education (1950).

(*f*) There appears to be a positive correlation between g and Youth. But great care must be taken in suggesting that g deterioration has set in in our groups by about 15 years of age or soon after. It is possible that our younger groups were really more intelligent than our older groups when the latter were 14½ years of age. But, if the younger groups were considerably more able than the older groups had been at an earlier age then a smaller negative correlation between Youth and *v* would have been expected.

Another factor analysis was carried out by the writer in which the *g*, *v* and *k* factors were first removed from the complete correlation matrix, and the remaining intercorrelations factor analysed using Thurstone's Centroid Method with subsequent rotation of axes to yield orthogonal group factors. This method generally confirmed the above results; as did the findings of two other analyses made independently by Prof. Vernon using twenty-three of the variables.

It is submitted then that strong evidence has been presented indicating that the ability to categorize is not adequately measured by current *g*, *v*, *k* tests; also that the ability is present and absent to some extent in non-clinical as in certain types of clinical cases. But two further points must be made:

(1) The variance contributed by the *c* factor is small. In other words, even in tests which have the highest loadings on the *c* factor, *g* contributed in most cases, even more to the test variance. When Sorting Tests are given individually, as seems usual in clinical work, it is possible that the tests are easier than those used in this experiment and thus the tests are less dependent on *g*. In such a case the deficit on the *c* factor would show up more clearly since the factor would account for more of the variability in scores.

(2) It is impossible to know at the moment whether the relative inability of our depressed group to categorize is due to the fact that the ability never developed, or whether it was present in greater measure and has since declined. Again, as has already been said, we cannot be sure that the deficit in the ability to categorize is due merely to the quality of the school teaching and cultural background generally, or whether it is, as Goldstein would have us believe, symptomatic of a change in the response of an individual to the environment.

In an attempt to test the hypothesis that the ability to categorize can be improved by intellectual stimulation a further group of young prisoners was tested. In this case each of the testees had served at least nine months in the prison during which time each prisoner had had to attend at least one evening class per week lasting 1½ hours and do one evening's Physical Training per week, quite apart, of course, from the normal work. There were also library facilities. Thirty such cases were matched against thirty prisoners in the first group for Matrices score, score on N.I.I.P. Group Test 70/23 and for Approved School and/or Borstal experience. Thus it was considered that we had thirty pairs matched for general intelligence. The scores obtained by the two groups on the Reading Test, Wechsler Vocabulary Test, Wisconsin Sorting Test, Vinacke Sorting Test, non-verbal Categorization Test, Proverbs Test and the Test of Three Activities were then studied, but the differences between the groups were not statistically significant except in the case of the non-verbal Categorization Test. On this test the second group of prisoners did rather worse than the first group. It is not possible to say at the moment, whether by intensive intellectual stimulation, the deficit could be made good.

Finally, it should be noted that our *c* factor contributes some of the variance of the Luchins type test. If this is confirmed by later research then it might be that the rigidity of Luchins is the same as the rigidity of Goldstein. As has already been said, however, the *c* factor is probably more complex than our analysis was able to show.

V. SUMMARY

1. There is an extensive literature surrounding the problem of intellectual deterioration in clinical cases. The results of past work are often difficult to interpret, due in the main to poorly designed experiments.

2. From an educational point of view we need to know if terms like 'conceptualization', 'rigidity of thinking' and so forth involve abilities other than *g*, *v*, *k* abilities.

3. Strong evidence has been presented indicating that there is an ability to group ideas or objects to a criterion and switch from one criterion to another not adequately measured by *g*, *v*, *k* tests.

4. Subjects in our non-stimulated group were unable to categorize as well as those of the stimulated group and in this sense behaved like Goldstein's brain lesion cases.

5. Whether a relative inability to categorize is due to poor-quality education and limited cultural background or whether it reflects a change in the response of an individual to the environment is a problem for further research.

6. The extent to which the discovery of our *c* factor will help in educational, occupational or clinical prediction is not yet known. The ability to categorize is probably more complex than our experiment was able to show.

The writer readily acknowledges the help that he received from Prof. P. E. Vernon in designing this experiment.

APPENDIX

Test	*g*	*v*	*k*	*c*	non-*p*	*n:ed*
1. N.I.I.P. 70/2	0·51	0·24	0·32	—	0·17	—
N.I.I.P. 70/3	0·54	0·20	0·25	—	—	—
2. Watts Vernon Reading	0·21	0·79	0·15	—	—	—
3. Arithmetic Problems	0·35	0·18	—	0·18	—	0·70
4. Wechsler Vocabulary	0·29	0·72	0·16	0·15	—	0·14
5L. Wisconsin Labels	0·58	0·31	—	0·26	—	—
5P. Wisconsin Pieces	0·56	0·30	—	0·28	—	—
6L. Vinacke Labels	0·51	0·22	—	0·43	0·15	—
6P. Vinacke Pieces	0·51	0·18	—	0·32	0·14	—
7A. Trist–Hargreaves Statements	0·66	0·26	—	0·18	—	0·22
7B. Trist–Hargreaves Pieces	0·58	0·17	—	0·34	—	0·17
8. Heibdreder Concept	0·37	0·46	—	0·23	—	0·25
9A. Non-verbal Categorization, Figures underlined	0·27	—	0·25	0·37	0·17	—
9B. Non-verbal Categorization, Reasons	0·44	—	0·24	0·44	—	—
10. Proverbs	0·39	0·61	—	—	0·26	—
11. Memory for Stories	0·30	0·56	0·26	—	0·19	—
12. Babcock Test of Memory for Sentences	0·21	0·44	—	0·16	—	—
13. Memory for Designs	0·29	—	0·62	—	−0·15	0·16
14. Creative Effort ZEK	0·15	—	0·20	0·14	0·53	—
15. Creative Effort 239	0·18	−0·20	—	—	0·57	—
16. Creative Effort VAT	—	−0·18	—	—	0·17	—
17. Plus and minus reversed	0·45	—	—	—	0·25	0·64
18. Multiplication and division reversed	0·37	—	—	—	0·14	0·75
19. Seasons reversed	0·38	0·15	—	0·16	—	0·40
20. Fractions	0·32	—	—	0·22	−0·21	0·18
21. Hidden Words	—	0·23	0·16	0·27	—	—
22. Beginnings and endings	−0·25	0·25	—	0·41	0·19	0·27
23. Three activities	—	0·16	—	0·29	0·34	0·16
24. Gottschaldt Figures	0·40	—	0·60	0·19	0·19	—
25. Shipley Abstraction	0·65	0·24	0·17	—	—	0·24
26. Kohs Blocks	0·45	—	0·62	0·19	—	—
Percentage variance	15·4	9·5	5·5	5·6	3·8	6·3

REFERENCES

BABCOCK, H. & LEVY, L. (1940). *Revision of the Babcock Examination for Measuring Efficiency of Mental Functioning.* Chicago: Stoelting.

BERG, E. A. (1948). A simple objective technique for measuring flexibility of thinking. *J. Genet. Psychol.* XXXIX, 15–22.

CATTELL, R. B. (1946). The riddle of perseveration. 'Creative Effort' and disposition rigidity. *J. Personality*, XIV, 229–267.

CROWN, S. (1951). Psychological changes following prefrontal leucotomy: a review. *J. ment. Sci.* XCVII, 49–83.

GOLDSTEIN, K. & SCHEERER, M. (1945). *Tests of Abstract and Concrete Thinking.* New York: Psychol. Corporation.

HALSTEAD, W. C. (1947). *Brain and Intelligence.* Chicago: University of Chicago Press.

HANFMANN, E. & KASANIN, J. (1937). A method for the study of concept formation. *J. Psychol.* III, 521–40.

HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.* XLII, 315–21.

HEIDBREDER, E. (1949). The attainment of concepts. VIII. The conceptualization of verbally indicated instances. *J. Psychol.* XXVII, 263–309.

HEY, C. G. (1951). M.A. Thesis, University of Birmingham Library.

KOUNIN, J. S. (1941). Experimental studies in rigidity. I. *Character & Pers.* IX, 251–72.

LEWIN, K. (1935). *A Dynamic Theory of Personality.* New York: McGraw Hill.

LUCHINS, A. S. (1942). Mechanization in problem solving. The effect of Einstellung. *Psychol. Monogr.* LIV.

LUCHINS, A. S. (1947). Proposed methods of studying rigidity of Behaviour. *J. Personality*, XV, 242–6.

MINISTRY OF EDUCATION (1950). *Reading Ability.* London: H.M.S.O.

NORRIS, K. E. (1940). *The Three R's and the Adult Worker.* McGill University.

SEMEONOFF, B. (1955). *A Manual of Diagnostic Performance Tests.* In preparation.

SHIPLEY, W. C. (1940). A self administering scale for measuring intellectual impairment and deterioration. *J. Psychol.* IX, 371–7.

VERNON, P. E. & PARRY, J. B. (1949). *Personnel Selection in the British Forces.* London: University of London Press.

VERNON, P. E. (1950). *The Structure of Human Abilities.* London: Methuen.

VINACKE, W. E. (1951). The investigation of concept formation. *Psychol. Bull.* XLVIII, 1–31.

WALKER, K. F., STAINES, R. G. & KENNA, J. C. (1943). *P* tests and the concept of mental inertia. *Character & Pers.* XII, 32–45.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins.

(*Manuscript received* 12 *July* 1954)

ADAPTATION TO LEVEL OF DIFFICULTY IN INTELLIGENCE TESTING

By A. W. HEIM

Psychological Laboratory, University of Cambridge

I. *Introduction* (pp. 211–212). II. *Test Material* (pp. 212–213). III. *Experimental design* (pp. 213–214). IV. *The subjects* (pp. 214–215). V. *Results* (pp. 215–220). VI. *Discussion* (pp. 220–221). VII. *Summary and conclusions* (pp. 221–222). VIII. *Acknowledgements* (p. 222). *Appendix* (p. 223). *References* (p. 224).

I. Introduction

There is some evidence that people unwittingly adapt their response to set tasks in accordance with the degree of difficulty presented. Naturally, individual variation obtains with respect both to this flexibility and to the type of response made. But in general people tend to 'rise to the occasion' in stimulatingly difficult circumstances and to 'sink to the occasion' in circumstances which present little or no challenge, giving then a mediocre performance on a relatively easy task. From this poor showing, the inference is sometimes drawn—perhaps invalidly—to a necessarily still poorer performance on a harder task.

Everyday life offers many illustrations of this type of adaptation. For example, players of team games are liable to excel themselves when confronted with formidable opponents and, conversely, to 'let the side down' when faced with a weak opposing team. The phenomenon has been displayed also in experimental situations. Mace (1931) demonstrated it in sensori-motor skills; Gibbs (1951) and others, in research on transfer of training; and Mowbray (1953) produced a very similar effect in his work on visual and auditory stimuli. Again, Heim's experiments (1949, 1950, 1954*b*) on repeatedly testing subjects with the same intelligence test suggested that the process has a part to play in mental testing. She found that a group of subjects of no more than average intelligence may produce a surprisingly high score if presented with a very high-grade intelligence test, believed on prior grounds to be 'too difficult' for them; whilst a group of the same intellectual level will fail to gain 100% on a much easier, and shorter, test of intelligence.

This suggests that the group tends to maintain a certain constancy of *relation* between performance and task, although the *absolute level* of performance may be less constant than is often supposed. It is clear that the phenomenon occurs only within certain limits: an impossible task, for instance, can scarcely evoke 'upward adaptation' from the group members who attempt it. On the other hand, the role of limits is not simple since, as is demonstrated in this paper and others, the limits themselves may shift as the subjects become upward- or downward-adapted (Heim, 1954*c*).

As already indicated, experiments on adaptation, in this sense, are numerous: though the terminology is far from uniform, studies on substantially the same problem are not lacking. But with the exception of the cited experiments on intelligence tests (and these were not primarily concerned with adaptation), the writer knows of no work in the field of mental testing which deals with adaptation to level of difficulty, either as the central issue or as a factor to take into account when interpreting results. On these grounds, the present investigation was initiated.

Intelligence test problems of known difficulty were incorporated into contexts of varying levels of difficulty; and these mixed tests were administered to groups of subjects previously equated on intelligence test performance. By these means it was hoped to shed some light on the extent and the mechanism of adaptation to level of difficulty, in the field of intelligence testing.

II. TEST MATERIAL

The test items used in these experiments were drawn from the second parts of tests AH 4 and AH 5 (Heim, 1947). On the basis of item analyses of these tests (Wallace, 1952; Watts, 1953), data were available concerning the relative difficulty of their individual problems. The thirty-six diagrammatic questions comprising part II of AH 5 were ranked in order from 1, the hardest (that is, the one yielding the smallest percentage of correct answers) to 36, the easiest. Nos. 1–12 were, arbitrarily, categorized as 'difficult AH 5'; nos. 13–24, as 'medium AH 5', and nos. 23–36, as 'easy AH 5'. The threefold classification was made in order to facilitate the task of producing four tests, differing in the number of AH 5 questions included, but roughly similar in respect to the difficulty of these AH 5 questions.

Four mixed tests (A 1, A 2, B 1 and B 2), each comprising forty items, were then constructed, using the 36 classified AH 5 questions and the easiest 44 AH 4 diagrammatic questions (all of which were easier than any of the AH 5 questions). These mixed tests will hereafter be referred to as 'AH 45 tests'. Selected equated groups were to take the two A tests and other equated groups were to take the two B tests. All the items in the A series appeared also in one or other of the B tests; but no A 1 items were common to A 2 nor were any B 1 items common to B 2, since one or other pair of tests was to be taken by the same subjects.

Table 1 shows the distribution of items among the four tests.

Table 1. *Distribution of AH* 4 *and AH* 5 *questions among tests A* 1, *B* 1, *B* 2 *and A* 2

		A 1		A 2			
		AH 5	AH 4	AH 5	AH 4	Total	
B 1	AH 5	—	—	14	—	14	40
	AH 4	—	16	—	10	26	
B 2	AH 5	6	—	16	—	22	40
	AH 4	—	18	—	—	18	
	Total	6	34	30	10		
		40		40			

Thus for example A 1, the easiest of the four tests, consisted of 34 AH 4 items and 6 AH 5 items; these 6 AH 5 and 18 of the AH 4 items appeared also in B 2, the remaining 16 AH 4 items appearing in B 1. It may be seen that the ascending order of difficulty of the four tests is as follows: I, A 1 (6 AH 5 items); II, B 1 (14 AH 5 items); III, B 2 (22 AH 5 items); IV, A 2 (30 AH 5 items).

By means of the threefold classification of AH 5 problems, these were, in accordance with their difficulty, divided among the four 'mixed' tests as evenly as possible. For instance, of the six AH 5 problems in A 1, two were 'difficult', two were 'medium' and two were 'easy'. When the four tests had had their items allocated, these were arranged cyclically so as to avoid consecutive series of AH 5 (or AH 4) questions, as far as possible. In test B 1, for instance, roughly every third item was an AH 5 question. The *type* of question too was distributed as evenly as was consistent with this plan, that is, a superimposition might be followed by an analogy and an analogy by a series—not another analogy. There was thus no gradient of difficulty in any of the four tests. The level of difficulty of adjacent problems varied markedly and, to all appearances, erratically. The only factor in common among all test questions was the form in which they were expressed: this was diagrammatic, the number of words used being at a minimum.

The decision was made to work with two pairs of tests rather than one pair, in order to ascertain whether the greater 'figure-ground' difference produced a greater effect, that is, whether adaptation to level of difficulty—in the event of its occurring at all—appeared more strikingly between tests A 1 and A 2 (the easiest and the hardest) than between (the intermediate) tests B 1 and B 2.

Time limits for the four tests were determined in accordance with the mean times which had been spent by university students on AH 4 and AH 5 questions, when taking the tests in unlimited time, as part of an earlier experiment (Heim, 1947). Eighty per cent of the mean time taken was allowed. When the time limits thus obtained were rounded off to the nearest minute, this method yielded the following set of time limits: A 1, 10 min.; B 1, 15 min.; B 2, 20 min.; and A 2 25 min.

III. EXPERIMENTAL DESIGN

Four groups of equated subjects were to take respectively A 1 followed by A 2, A 2 followed by A 1, B 1 followed by B 2 or B 2 followed by B 1. Thus the total testing time for all subjects on their pair of mixed tests was to be 35 min. Preliminary examples were, regretfully, dispensed with, as they might well have produced some unmeasurable adaptation to level of difficulty.

Some prior testing was necessary, however, since it was essential to equate the four groups on intelligence test performance. Thus all the subjects had in fact had previous experience on at least one other test, which was somewhat similar to the AH 45 series. The test selected for this purpose was Penrose's Pattern Perception (1944, 1946). This was chosen because it is a high-grade test intended for use with subjects of university status and it also has a diagrammatic bias. It comprises eight preliminary examples and sixty-four test items. These are classificatory in kind, the subject in every case being presented with five diagrams and required to select the 'odd man out'. A time limit of 30 min. (exclusive of the examples) is imposed.

On the basis of their Pattern Perception scores, the total group was to be divided into four subgroups with similar means and variances. The time interval between the Pattern Perception and the AH 45 testing sessions ranged from 4 to 7 days. Subjects were naturally not informed of the purpose of the inquiry until they had completed both testing sessions. They were asked to take part in an experiment which, it was hoped, would yield information on some new tests and they were told that their presence on both occasions was necessary if their participation was to be of value.

At the second session, the test room was divided into four sections, each subject being directed to his own section as he entered the room. All subgroups were started on their first test at the same time and were asked to stop working on their second test at the same time. After 10 min., the A 1's had their papers removed and were presented with A 2 tests, which they began as soon as they had written their name on the cover; the B 1's were similarly changed to B 2's after 15 min; and so on. The groups were told at the beginning how long they would have for their first test but were given no information about the time limit on their second test. The complete schedule is shown in the Appendix, Table I.

It was hoped that by designing the experiment in this way, data would be gained on both immediate and successive adaptation to level of difficulty. If the hypothesis of *immediate* adaptation were confirmed, then the AH 5 problems common to an easy and a difficult test would tend, on first AH 45 testing, to be answered more correctly in the difficult than in the easy context. If *successive* adaptation occurred, that is, between two tests of different levels taken consecutively, then subgroups passing from an easy to a harder test would produce relatively low scores on comparable questions in their second test; whereas subjects beginning with their more difficult test would be likely to score relatively highly on the crucial questions in their second test.

The prediction with regard to AH 4 questions was less clear. These questions were included as creators of an easy context rather than for their own intrinsic interest. It was thought possible, however, that the response to them too might be affected by *their* context; that even these simple problems might yield results determined in part by the difficulty of the problems preceding and succeeding them.

Equated rather than identical groups were used, in order to obviate the confusion of practice effects with adaptation effects. As will be seen, the results on second testing did indicate effects on *speed* of work, which seemed to be due to a combination of practice and of differing time limits.

IV. THE SUBJECTS

The experiments reported in this paper were conducted in the U.S.A. A pilot experiment on similar lines, carried out by the writer before her visit to America, suggested that the hypothesized adaptation did take place and that it might occur more markedly in highly intelligent than in less intelligent groups (at least with the high-grade tests which she was using). She was anxious therefore to conduct inquiries with American groups of differing intellectual levels.

Three groups were chosen: graduates of research status at a private university in the West; undergraduates, mostly in their second year, at the same university; and men and women 'basic entrants' to an Air Force base in the South. The graduates were all volunteers; the majority were scientists but they included also historians, educationists and mathematicians. Their average age was just under 29 years. Of the 61, 47 were men; the 14 women were distributed as evenly as possible among the four subgroups.

119 undergraduates took part in the inquiry. Many of these were majoring in Psychology and these participated largely in order to make up their requisite number of 'experimental' hours. A substantial minority, however, were volunteers majoring in a wide variety of subjects. The average age of this group was 19 years. Here the women preponderated over the men, 71 to 48.

The 114 American Air Force subjects comprised 58 male and 56 female 'basic entrants'. These were not volunteers: the tests were included as part of their first fortnight's training, which consisted mainly of lectures and of tests, some 'operational' and some (such as those reported on here) 'experimental'. The average age of this group was twenty years.

The testing conditions for these subjects were far less satisfactory than were those of the university groups. For example, the Air Force test groups were bigger and the supervision less adequate; many of the subjects were obviously lacking in sleep (it transpired later that some had been up since 4·00 and even 3·00 a.m.) and, understandably, were lacking in interest and incentive.

Those university subjects who were interested had the purpose of the inquiry explained to them after they had all completed their second session. There was no opportunity for such explanation with the Air Force subjects. The writer merely requested their help (which they knew they were not entitled to refuse) in 'trying out a new series of tests' and she did not have the opportunity of seeing them again after the data had been obtained.

V. RESULTS

Table 2 shows the mean scores gained by the university subgroups and the Air Force subgroups on their Pattern Perception test, first AH 45 and second AH 45 testing. It may be seen that the Pattern Perception means and standard deviations for both sets of subgroups are sufficiently similar to justify treatment of these subgroups as identical. None of these intra-subgroup differences approaches statistical significance. The figures for the graduates and undergraduates in separate groups are shown in the Appendix, Table II. Since the differences in test score between the graduates and the undergraduates were so very slight, it was thought best to amalgamate them into one large 'university group'. It may be seen that combining the two groups had the further advantage of rendering still more homogeneous the standard deviations of the subgroups.

Table 2. *Pattern Perception and AH* 45 *test results*

			AH 45			
	Pattern perception		First testing		Second testing	
n	Mean	S.D.	Test	Mean	Test	Mean
			University			
48	48·96	6·60	A 1	30·98	A 2	24·23
45	49·69	6·88	B 1	28·98	B 2	27·09
43	49·67	6·18	B 2	27·00	B 1	29·30
44	49·59	7·28	A 2	21·95	A 1	25·32
			Air Force			
30	38·23	8·09	A 1	20·63	A 2	16·00
26	37·58	8·72	B 1	20·04	B 2	19·69
28	38·07	8·32	B 2	17·21	B 1	18·61
30	37·90	7·53	A 2	16·60	A 1	17·67

If we examine the mean AH 45 scores on first testing, we find the following order of difficulty for the two main groups: A 1 (easiest), B 1, B 2, A 2 (hardest). This is the 'objectively correct' order of difficulty for the four tests, judged by the respective numbers of AH 5 questions. Evidently the more generous time limits allowed for tests A 2 and B 2 do not compensate for the higher proportions of harder questions which they contain. The trend would presumably have been the same, with a steeper slope, had a uniform time limit been imposed for the four tests.

In Table 3 are shown the mean numbers of questions attempted by the two main groups, on first and second testing, respectively. An 'attempted' question is one which has been answered, correctly or incorrectly, or which has been omitted, at least one later question having been answered. An omitted question was classed as 'attempted' since the subject had evidently considered it at least long enough to decide to refuse it before going on to the next item. The subjects were asked to work through the problems in order, and most of them appeared to follow this instruction. Thus the figures in Table 3 probably reflect fairly accurately differences in rate of work.

Table 3. *Mean number of questions attempted*

		University group		Air Force group	
Test	Time limit (min.)	First testing	Second testing	First testing	Second testing
A 1	10	35·7	28·3	28·4	23·5
B 1	15	36·7	35·9	32·3	32·6
B 2	20	36·2	37·1	33·6	35·6
A 2	25	32·5	36·1	36·8	37·8

The interesting point about Table 3 is the influence that the time limit of the first testing in every case exerted on the subject's speed in his second testing (see arrows). The short time limits of A 1 and B 1 produced an *increased* rate of work in the second session; the relatively long time limits of B 2 and A 2 produced a *slower* rate of work in the second session. It may be seen that this effect was quite consistent, though it was more marked for some groups and some pairs of tests than for others.

This result may be seen more clearly if instead of mentally following the *subgroups* from their first to their second test, we consider the *tests*, taken first by one subgroup and then by another, equated, subgroup. Thus we see that A 1 in the university group, for instance, is tackled far more slowly (mean number attempted = 28·3) when it is taken after A 2 than when it is taken first (mean number attempted = 35·7); whereas A 2 is tackled more quickly when it is taken second (mean number attempted = 36·1) than when taken first (mean number attempted = 32·5). This result holds in every case except the Air Force group on B 1.

The writer considers this finding important for two reasons. First, it illustrates once again (Heim, 1954*a*, *b*) the artificial nature of designating particular tests as speed tests and particular groups as fast or slow, *in vacuo*, regardless of instructions and of the total test situation. Secondly, it provides some evidence for the somewhat complicated interpretation given below of the results obtained from the second testings in this inquiry.

It may be seen in Table 2 that on second, unlike on first, testing the means of the four tests follow no obvious pattern. The orders of difficulty for the Air Force and the university groups show no particular trend nor do they agree. They have in common only A 2 as the lowest mean, on second testing (as on first).

The combined data in Tables 3–5, and in the completed answer sheets, suggest that the lack of definite pattern in the second testing scores may be due to a conflict between two tendencies. On the one hand, there is some practice effect from the immediately preceding first testing, as described above: this results primarily in a slowing down (with consequent decreased opportunity for scoring) of those subjects who took first a harder test, with its longer time limit; and a speeding up (with increased opportunity for scoring) of those who took first their easier test, with its shorter time limit. On the other hand, there is some successive adaptation to level of difficulty: this results in augmented power, or accuracy, *per question*, on the part of those subjects who took first their harder test and diminished power, or accuracy, per question, for those who took first their easier test.

This successive adaptation in turn conflicts with immediate adaptation. These latter points emerge only on examination of the mean scores per question, as shown in Tables 4 and 5. The mean in each case has been calculated by dividing the number of correct answers to each question by the number of subjects who attempted the question—not by the total number of subjects in the group. This method was found necessary in order to allow for the considerable variation in the number of questions attempted among the members of each group. The figures showing the means based on total number of subjects per subgroup are given in the Appendix, Tables III and IV.

Table 4. *Mean scores on classified test questions*[1]

n		Test	AH 45 questions		AH 5 questions		AH 4 questions	
			No. of questions	Mean	No. of questions	Mean	No. of questions	Mean
				University				
48	First testing	A 1	40	0·87	6	0·50	34	0·94
45		B 1	40	0·79	14	0·51	26	0·94
43		B 2	40	0·75	22	0·58	18	0·96
44		A 2	40	0·68	30	0·59	10	0·94
44	Second testing	A 1	40	0·89	6	0·58	34	0·95
43		B 1	40	0·82	14	0·59	26	0·94
45		B 2	40	0·73	22	0·55	18	0·95
48		A 2	40	0·67	30	0·58	10	0·95
				Air Force				
30	First testing	A 1	40	0·73	6	0·35	34	0·79
26		B 1	40	0·62	14	0·37	26	0·76
28		B 2	40	0·51	22	0·31	18	0·76
30		A 2	40	0·45	30	0·32	10	0·85
30	Second testing	A 1	40	0·75	6	0·30	34	0·83
28		B 1	40	0·57	14	0·28	26	0·74
26		B 2	40	0·55	22	0·34	18	0·82
30		A 2	40	0·42	30	0·31	10	0·77

[1] Means based on number of questions attempted

Let us first consider immediate adaptation to level of difficulty. This entails inspecting the results on first testing, shown in Table 4. The order of difficulty of the four tests, as shown by the mean score per question gained by each subgroup, is, as mentioned above, evidently determined by the number of AH 5 items included in each of the tests (yielding the order A 1, B 1, B 2, A 2). When, however, we turn to the corresponding mean scores for AH 5 questions alone on first testing, we find that for the university group this order is reversed. It is now as follows: A 2 (highest), B 2, B 1 and A 1 (lowest mean score).

In order to establish the relationship between the mean AH 5 score and the number of AH 5 items present in a test, the linear regression of the former on the latter was estimated. This was found to have a slope significantly different from 0: $P = 0{\cdot}05$ using a one-tailed test.

It is suggested that this result is due to the adaptation of subjects to the level of difficulty of the test with which they are presented. For example, subjects taking A 1 as their first test have only six AH 5 questions among their forty: they therefore gain a higher overall score (0·87) than do members of the other three subgroups. On the other hand, they become attuned or adapted, as they work through the test, to a lower level of difficulty and therefore, despite their highest overall score, they obtain the lowest score of the four groups on AH 5 items (0·50). The position is exactly reversed for the subjects taking A 2 first with its high proportion of AH 5 items (thirty out of forty): this group gains the lowest overall score (0·68) and yet achieves the highest AH 5 score (0·59).

It will be remembered (Table 1) that every question appeared in two tests, once in A 1 or A 2 and again in B 1 or B 2. Let us consider the AH 5 questions from this point of view. On the hypothesis of upward adaptation, it should be found that when identical AH 5 questions are presented (to equated groups) in two tests, whose other items differ in level of difficulty, the AH 5 questions tend to be answered more correctly when they appear in the harder context than when they appear in the easier context.

Table 5. *Mean scores on questions common to two tests*[1]

	AH 5 questions in common					AH 4 questions in common				
	No. of questions	Mean				No. of questions	Mean			
		A 1	B 1	B 2	A 2		A 1	B 1	B 2	A 2
				University						
First testing	—	—	—	—	—	16	0·94	0·94	—	—
	6	0·50	—	0·66	—	18	0·93	—	0·96	—
	14	—	0·51	—	0·63	10	—	0·95	—	0·94
	16	—	—	0·55	0·55	—	—	—	—	—
Second testing	—	—	—	—	—	16	0·94	0·93	—	—
	6	0·58	—	0·58	—	18	0·96	—	0·95	—
	14	—	0·59	—	0·60	10	—	0·96	—	0·95
	16	—	—	0·54	0·56	—	—	—	—	—
				Air Force						
First testing	—	—	—	—	—	16	0·76	0·74	—	—
	6	0·35	—	0·34	—	18	0·82	—	0·76	—
	14	—	0·37	—	0·34	10	—	0·79	—	0·85
	16	—	—	0·30	0·30	—	—	—	—	—
Second testing	—	—	—	—	—	16	0·82	0·73	—	—
	6	0·30	—	0·40	—	18	0·84	—	0·82	—
	14	—	0·28	—	0·29	10	—	0·74	—	0·77
	16	—	—	0·32	0·32	—	—	—	—	—

[1] Means based on number of questions attempted.

This does in fact occur with the university group (see Table 5), on first testing. Examination of the three pairs of tests possessing AH 5 questions in common, reveals that the harder test of the pair produces a higher mean score on those questions than the easier test of the pair, in all cases but one. In order to test the significance of the differences among the member-questions of the pairs of AH 5 means, a count was made of the number of times the increase was in the predicted direction. On the null hypothesis this should be about half of them; it was therefore possible to use a (one-tailed) binomial test of significance. In the case of A 1 (easiest test) and B 2 (second hardest), six of the six questions showed the predicted increase ($P=0{\cdot}016$); in the case of B 1 (second easiest) and A 2 (hardest test), twelve of the fourteen questions showed it ($P=0{\cdot}007$).

The pair of tests which does not yield this result is the B 2–A 2 pair. The question by question analysis here revealed no significant difference, only seven out of the sixteen questions showing the predicted increase in score. It may be seen in Table 5 that the AH 5 items in common did, in fact, produce identical mean scores (0·55) in the two contexts. A 2 and B 2 are the most difficult two tests and are the only pair of tests of 'adjacent' difficulty to be compared (since A 1 and B 1 have no AH 5 questions in common). This finding is, therefore, not surprising.

The trend of university results on AH 5 questions on second testing is similar but less striking. Again two of the three pairs of tests show a difference in the predicted direction, and the tests comprising the third pair produce identical mean scores. But the differences on second testings are very slight, probably owing to the differences in speed of work induced by the differing time limits for first testing and to *successive* adaptation to level of difficulty.

The subject whose first test was difficult (i.e. A 2 or B 2) will have adapted to a high level. He will therefore embark on his second (easier) test in an 'upwardly adapted' state, doing better on the AH 5 items than his equated fellows did when taking the same easy test first. This phenomenon shows clearly if we compare the respective AH 5 mean scores gained by university subjects on first and second testing (see Table 4). We are thus comparing a test taken on one occasion as a group's first test with the same test taken on another occasion as (an equated group's) second test.

The differences are not great but, in all four pairs, they go in the direction to be expected on the hypothesis of successive adaptation. A 1 and B 1, the easier two tests, respectively produce a *higher* mean score on their AH 5 questions when taken second, that is, when the subjects are upwardly adapted, having just taken a difficult test. Whereas A 2 and B 2, the harder two tests, respectively produce a *lower* mean AH 5 score when taken second, that is when the subjects are downwardly adapted, having just taken an easy test. By means of counting the number of times this occurred in individual questions it was possible to test the significance of these findings (see Table 6).

Table 6

Test	No. of AH 5 questions	No. showing predicted *increase*	P
A 1	6	6	0·016
B 1	14	11	0·029
		No. showing predicted *decrease*	
B 2	22	15	0·07
A 2	30	16	—

Examination of the university AH 4 results, however, yields little of interest. These questions proved so easy for the graduates and undergraduates that the mean score per question on number attempted for every subgroup lies between 0·94 and 0·96 for both first and second testing, rendering meaningless any examination for trends or for differences between means. It can only be said that for the university group, the AH 4 items provided the requisite easy background.

The Air Force results are markedly different. Their range on AH 4 items exceeded that on AH 5 items (being 0·74–0·85 and 0·28–0·37, respectively). Interestingly, on first testing, these subjects obtain their highest AH 4 mean in test A 2, the hardest test, suggesting that AH 4 may be to the Air Force as AH 5 is to the university. But inspection of the AH 4 questions in common to pairs of tests, and of the AH 4 mean scores on second testing, reveals no interpretable pattern.

It may be seen that almost all the Air Force results on AH 5 lie between 0·3 and 0·4. Thus these subjects were answering about a third of their questions correctly. This is approaching random answering, which would yield one in five. This suggests that AH 5 is almost outside the range of most of the Air Force subjects—at least in the state of fatigue and poor motivation in which they took the test.

There are no significant trends of differences among the Air Force AH 5 scores whether we examine the score-order of the AH 5 questions on first testing or the scores gained on AH 5 questions in common. If anything, the differences on these latter questions go the 'wrong' way, the slightly higher scores being gained on the AH 5 questions in their *easier* contexts. There is thus no evidence of immediate adaptation in the Air Force results. Similarly, no sign of successive adaptation is revealed on inspecting the scores on AH 5 questions taken on first and second testing. The Air Force results reveal no consistent pattern of any kind, on either the easier or the harder test questions.

VI. Discussion

A small-scale pilot experiment on the same lines, carried out with English subjects, had yielded similar results. In this case, the two groups had been respectively university students and naval ratings. In this earlier experiment, too, the more intelligent, more highly educated, volunteer subjects had adapted to level of difficulty, as defined above, whereas the less intelligent, less educated, conscripted subjects gave no evidence of such adaptation.

It is not clear whether the failure to 'adapt upward' is due mainly to the poorer motivation and/or level of intellect of the non-volunteering groups or to the *relationship* between their level and that of the tests. Previous inquiries on related problems have demonstrated the importance of this latter point (Heim, Wallace & Cane, 1949, 1950; Heim, 1954*b*). It would be interesting to repeat the experiment using AH 4 questions and still easier questions, with two groups comprising, say, naval ratings and another, duller, group. The problem of volunteers versus conscripts in this connexion also merits investigation. The writer's impression is that the naval ratings would in such a situation respond to AH 4 questions in their easy context as the university groups responded to AH 5 questions in their relatively easy context.

The evidence so far is not conclusive. The results do tally, however, with observations made both outside the psychological laboratory and in non-psychometric experimental situations (Mace, 1931; Gibbs, 1951; Mowbray, 1953; Helson, 1948). Many of the factors determining behaviour in real life situations have traditionally been ignored in the group mental testing situation. But there is no reason to assume that they do not operate; indeed, such factors may play a role in producing some of the errors and discrepancies which have been grouped together, and corrected for, under the blanket term 'unreliability'.

If it be true that some subjects unwittingly adapt upwards when faced with test problems presenting them with a serious challenge and that the same subjects adapt downwards, failing to gain 100 % when tackling far easier problems than those they have shown they can solve, then inaccuracies in the assessment of intelligence of individuals, and even of groups, may sometimes arise. Presumably this is especially likely to occur, unrecognized, when all the test problems set are of roughly the same, inappropriate, level. Such questions as the gradient of difficulty, form of test instructions and methods of scoring (weighted and unweighted) immediately present themselves as demanding of more respect than they have been accorded in the past.

It would follow that the notion of one intelligence scale along which the total population can be distributed is untenable. Such a notion implies that every individual possesses a constant amount of a static quality, measurable by standard tests of intelligence. In the writer's opinion there are other objections to this view, many of which are not easily susceptible of experimentation. But the findings of this paper, if convincing and if confirmed by other investigators, would suggest that the subject's response to the same test problems will vary; he will sometimes do himself less than justice (still on the *test* criterion) and he will sometimes do 'surprisingly well'. Experimental evidence has been adduced in favour of one possible reason for such 'test unreliability'.

Thus, a number of practical considerations would follow on acceptance of the phenomenon of adaptation to level of difficulty in mental testing. It would impose, for instance, far greater care than is usually exercised in the earlier stages of testing: before administering an intelligence test to a group, the intellectual level and homogeneity of the group, and the level of the test, would be considered in relation to one another. The aim would be to allow scope for the maximum possible upward adaptation to level of difficulty, that is, the test chosen would be as difficult for the subjects as is consistent with their understanding the problems set. This suggestion is made in the belief that it is more valuable when testing to determine the subject's upper limits than to determine—albeit unreliably—what he *cannot* do.

The phenomenon would have not only practical consequences. If proved conclusively, it might have some effect on the theory of intelligence and intelligence testing since it would throw doubt on the existence of a static entity measurable and expressible in purely quantitative terms, and isolable from other characteristics, notably those usually subsumed under 'personality'.

Limitations of space preclude a full discussion here. It seems evident, however, that if the level of difficulty of a problem with which an individual can cope is partly dependent on the level of difficulty of the context, generalizations about the subject's ability based on problem-solving, without reference to the present or the immediately preceding context, are exceedingly risky; and generalizations about 'objectively based' levels of difficulty of particular problems *in vacuo* have little worth.

VII. SUMMARY AND CONCLUSIONS

1. The hypothesis was advanced that test subjects tend, unwittingly, to adapt to the level of difficulty of their test items. Upward adaptation was hypothesized when the test, whilst not grossly exceeding the limits of the subjects' ability, was sufficiently hard to extend them fully; and downward adaptation was hypothesized when the test was correspondingly simple. Previous experimenters have reported this phenomenon in non-psychometric situations.

2. In an attempt to verify this hypothesis in the field of mental testing, two pairs of tests were prepared using a selection of items from one high-grade and one easier intelligence test. The four tests comprised the same number of questions but they differed in their number of 'difficult' problems and in their time limits, the harder tests being allotted proportionately longer time limits. A group of subjects was selected, given a

preliminary, standard test of intelligence and, on the basis of their performance, divided into four equated subgroups. Each of these subgroups took one or other pair of 'mixed' tests in one or other order. Thus identical problems were presented, in contexts of varying degrees of difficulty, to equated subjects.

3. The experiments were carried out in the U.S.A., the two main groups who provided the data consisting of: *A*, graduates and undergraduates at a private university in the West; *B*, basic entrants at a U.S. Air Force base in the South.

4. The main results yielded by group *A* were as follows:

(i) Evidence of *immediate* adaptation to level of difficulty was found when a comparison was made of the scores gained on difficult questions in an easy context and a difficult context, respectively. The mean score on the difficult items was found to bear a statistically significant relationship to the proportion of difficult to easy items in the test: the more numerous the harder questions, the higher their mean score (p. 217). Moreover, identical difficult problems were solved more often (by equated subjects) when they appeared in their harder context (p. 218).

(ii) The data suggested also the operation of *successive* adaptation to level of difficulty. Subjects taking first the easier test of the pair gave evidence of some downward adaptation, achieving fewer correct solutions to the difficult problems in their (harder) second session than did members of the equated group when taking this (hard) test for the first time. Similarly, upward adaptation was demonstrated by those who took first their more difficult test: these subjects showed greater accuracy per question on the difficult problems in their (easier) second session than did equated subjects taking this (easy) test first (p. 219).

(iii) The time limit imposed on the test taken first appeared to influence the speed with which subjects tackled their second test. Those whose first test was easy, and therefore short, worked faster, attempting a greater number of items, in their second test (although it was more difficult); whereas those whose first test was hard, and therefore long, worked more slowly on their second test (although it was easier) (p. 216).

5. The data yielded by group *B*, the Air Force subjects, revealed no interpretable pattern. Their responses to the difficult questions approached sufficiently closely to random answering to suggest that with subjects of such intellectual level, tests of a lower standard should be used.

6. A small-scale pilot experiment previously conducted in England with the same two pairs of tests on a group of university students and a group of naval ratings, had produced substantially similar results.

7. It is suggested that the negative findings of the duller groups are due to the harder test problems being virtually outside their range, and that it would be worth while repeating the investigation with such groups using lower grade tests.

8. Some of the implications of the phenomenon of adaptation to level of difficulty in mental testing are briefly discussed.

VIII. ACKNOWLEDGEMENTS

The writer's thanks are due to the administrators of the Smith-Mundt and Fulbright Grants and to the Medical Research Council for financial support; to the Director of the Human Resources Research Centre at the Air Force base and of Stanford University

Psychology Department for making available the test groups and testing facilities; to the subjects for providing the data; to Prof. L. S. Penrose for allowing the use of his Pattern Perception Test; to Mr J. W. Whitfield for his help in designing the experiment; and to Miss Violet Cane, Mr E. G. Chambers and Miss Kathleen Watts for their criticism and their advice on the statistical analysis of the results.

APPENDIX

Table I. *Schedule of testing*

Subgroup	1st testing immediately followed by 2nd testing		Total time
i	A 1 (10 min.)	A 2 (25 min.)	35 min.
ii	B 1 (15 min.)	B 2 (20 min.)	35 min.
iii	B 2 (20 min.)	B 1 (15 min.)	35 min.
iv	A 2 (25 min.)	A 1 (10 min.)	35 min.

Table II. *Pattern Perception and AH 45 test results*

	Pattern perception		AH 45			
			First testing		Second testing	
n	Mean	S.D.	Test	Mean	Test	Mean
			Graduates			
16	49·75	7·19	A 1	33·62	A 2	29·06
15	49·27	7·93	B 1	27·40	B 2	28·20
15	49·75	8·12	B 2	27·00	B 1	28·20
15	49·60	7·36	A 2	24·60	A 1	26·33
			Undergraduates			
32	48·56	6·36	A 1	29·66	A 2	21·81
30	49·90	6·43	B 1	29·77	B 2	26·53
28	49·75	5·01	B 2	27·00	B 1	29·89
29	49·59	7·37	A 2	20·59	A 1	24·79

Table III. *Mean scores on classified test questions*[1]

			AH 45 questions		AH 5 questions		AH 4 questions	
n		Test	No. of questions	Mean	No. of questions	Mean	No. of questions	Mean
				University				
48	First testing	A 1	40	0·77	6	0·45	34	0·83
45		B 1	40	0·72	14	0·48	26	0·86
43		B 2	40	0·68	22	0·53	18	0·85
44		A 2	40	0·55	30	0·48	10	0·75
44	Second testing	A 1	40	0·63	6	0·41	34	0·67
43		B 1	40	0·73	14	0·54	26	0·84
45		B 2	40	0·68	22	0·52	18	0·87
48		A 2	40	0·61	30	0·52	10	0·85
				Air Force				
30	First testing	A 1	40	0·52	6	0·26	34	0·56
26		B 1	40	0·50	14	0·30	26	0·61
28		B 2	40	0·43	22	0·27	18	0·63
30		A 2	40	0·42	30	0·30	10	0·77
30	Second testing	A 1	40	0·44	6	0·18	34	0·49
28		B 1	40	0·46	14	0·23	26	0·59
26		B 2	40	0·49	22	0·30	18	0·73
30		A 2	40	0·40	30	0·29	10	0·72

---

[1] Means based on number of questions in the test.

Table IV. *Mean scores on questions common to two tests*[1]

	AH 5 questions in common					AH 4 questions in common				
	No. of questions	Mean				No. of questions	Mean			
		A 1	B 1	B 2	A 2		A 1	B 1	B 2	A 2
				University						
First testing	—	—	—	—	—	16	0·86	0·83	—	—
	6	0·45	—	0·63	—	18	0·81	—	0·85	—
	14	—	0·48	—	0·53	10	—	0·90	—	0·75
	16	—	—	0·49	0·44	—	—	—	—	—
Second testing	—	—	—	—	—	16	0·68	0·80	—	—
	6	0·41	—	0·56	—	18	0·66	—	0·87	—
	14	—	0·54	—	0·55	10	—	0·89	—	0·85
	16	—	—	0·50	0·50	—	—	—	—	—
				Air Force						
First testing	—	—	—	—	—	16	0·55	0·55	—	—
	6	0·26	—	0·29	—	18	0·57	—	0·63	—
	14	—	0·30	—	0·32	10	—	0·70	—	0·77
	16	—	—	0·26	0·28	—	—	—	—	—
Second testing	—	—	—	—	—	16	0·47	0·55	—	—
	6	0·18	—	0·36	—	18	0·50	—	0·73	—
	14	—	0·23	—	0·28	10	—	0·65	—	0·72
	16	—	—	0·28	0·30	—	—	—	—	—

[1] Means based on number of questions in the test.

REFERENCES

GIBBS, C. B. (1951). Transfer of training and skill assumptions in tracking tasks. *Quart. J. Exp. Psychol.* III, 99–110.

HEIM, A. W. (1947). An attempt to test high grade intelligence. *Brit. J. Psychol.* XXXVII, 70–81.

HEIM, A. W., WALLACE, J. G. & CANE, V. R. (1949, 1950). The effects of repeatedly retesting the same group on the same intelligence test. Parts I, II and III. *Quart. J. Exp. Psychol.* I, 151–9; II, 19–32 and 182–97.

HEIM, A. W. (1954*a*). *The Appraisal of Intelligence*, chap. IX. London: Methuen.

HEIM, A. W. (1954*b*). Les conventions à la base des tests d'intelligence. *Enfance*, 105–12.

HEIM, A. W. (1954*c*). Psychological adaptation as a response to variations in difficulty and intensity. Unpublished.

HELSON, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.* LV, 297–313.

MACE, C. A. (1931). The influence of indirect incentives upon the accuracy of skilled movements. *Brit. J. Psychol.* XXII, 101–14.

MOWBRAY, G. H. (1953). Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty. *J. Exp. Psychol.* XLVI, 365–72.

PENROSE, L. S. (1944). An economical method of presenting matrix intelligence tests. *Brit. J. Med. Psychol.* XX, 144–6.

PENROSE, L. S. (1946). The Pattern Perception Test: experimental results. *Brit. J. Med. Psychol.* XX, 339–43.

WALLACE, J. G. (1952). The results of a high grade intelligence test applied to a university population. *Brit. J. Psychol.* XLIII, 61–9.

WATTS, K. P. (1953). Influences affecting the results of a test of high grade intelligence. *Brit. J. Psychol.* XLIV, 359–67.

(*Manuscript received* 30 *July* 1954)

A TEST OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEANS OF CORRELATED MEASURES, BASED ON A SIMPLIFICATION OF STUDENT'S *t*

By JOSEPH SANDLER

The Tavistock Clinic, London

Psychologists have often to compare the means of two distinct sets of scores, where there is a correlation between the two sets. This is usually done when:

(i) Two groups have been matched with respect to some extraneous variable, such as age or intelligence. Each person in one group is paired with a person in the other.

(ii) A set of persons has been given a test twice, in different circumstances (e.g. before and after training, or the administration of a drug).

The appropriate test of significance of the difference between the two means is Student's t by the method of differences (or by the more laborious method in which the correlation between the two sets of scores is actually calculated and then corrected for). t by the method of differences is given as follows (Johnson, 1949):

$$t=\frac{\bar{d}}{\text{S.E.}_{\bar{d}}},$$

where $\bar{d}$ is the mean difference between the pairs of scores, and $\text{S.E.}_{\bar{d}}$ is the standard error of the mean difference. t has $N-1$ degrees of freedom, where N is the number of cases in each group.

It is the purpose of this paper to present a statistic A, which yields exactly the same results as Student's t, being rigorously derived from it, but which is extremely simple to calculate. If, as is commonly the case, two matched groups are compared with respect to a large number of variables, the use of A results in a considerable saving of time and labour. It is found as follows:

$$A=\frac{\text{The sum of the squares of the differences}}{\text{The square of the sum of the differences}}=\frac{\Sigma d^2}{(\Sigma d)^2}.$$

A table of A has been constructed, and is appended. The table is entered with $N-1$ degrees of freedom, and a value of A is significant at any given level if it is equal to or *less* than the tabled value. It corresponds to the two-tail form of t.

If we write t as

$$t=\frac{(\Sigma d)/N}{\sqrt{\dfrac{\Sigma d^2-\dfrac{(\Sigma d)^2}{N}}{N(N-1)}}},$$

taking the square of both sides gives us

$$t^2=\frac{N-1}{N}\,\frac{(\Sigma d)^2}{\Sigma d^2-\dfrac{(\Sigma d)^2}{N}}.$$

Hence,
$$\frac{N-1}{Nt^2}=\frac{\Sigma d^2}{(\Sigma d)^2}-\frac{1}{N}.$$

From this we may write A in terms of t:
$$A=\frac{N-1}{Nt^2}+\frac{1}{N},$$

t can be written in terms of A as follows:
$$t=\sqrt{\frac{N-1}{AN-1}}.$$

The table of A has been constructed from published tables of Student's t (Fisher & Yates, 1948) and the tables of the inverted beta (F) distribution, computed by Merrington & Thompson (1943).

Table of A

(For any given value of $N-1$, the table shows the values of A corresponding to various levels of probability. A is significant at a given level if it is equal to or *less than* the value shown in the table.)

	Probability					
$N-1$	0·10	0·05	0·02	0·01	0·001	$N-1$
1	0·5125	0·5031	0·50049	0·50012	0·5000012	1
2	0·412	0·369	0·347	0·340	0·334	2
3	0·385	0·324	0·286	0·272	0·254	3
4	0·376	0·304	0·257	0·238	0·211	4
5	0·372	0·293	0·240	0·218	0·184	5
6	0·370	0·286	0·230	0·205	0·167	6
7	0·369	0·281	0·222	0·196	0·155	7
8	0·368	0·278	0·217	0·190	0·146	8
9	0·368	0·276	0·213	0·185	0·139	9
10	0·368	0·274	0·210	0·181	0·134	10
11	0·368	0·273	0·207	0·178	0·130	11
12	0·368	0·271	0·205	0·176	0·126	12
13	0·368	0·270	0·204	0·174	0·124	13
14	0·368	0·270	0·202	0·172	0·121	14
15	0·368	0·269	0·201	0·170	0·119	15
16	0·368	0·268	0·200	0·169	0·117	16
17	0·368	0·268	0·199	0·168	0·116	17
18	0·368	0·267	0·198	0·167	0·114	18
19	0·368	0·267	0·197	0·166	0·113	19
20	0·368	0·266	0·197	0·165	0·112	20
21	0·368	0·266	0·196	0·165	0·111	21
22	0·368	0·266	0·196	0·164	0·110	22
23	0·368	0·266	0·195	0·163	0·109	23
24	0·368	0·265	0·195	0·163	0·108	24
25	0·368	0·265	0·194	0·162	0·108	25
26	0·368	0·265	0·194	0·162	0·107	26
27	0·368	0·265	0·193	0·161	0·107	27
28	0·368	0·265	0·193	0·161	0·106	28
29	0·368	0·264	0·193	0·161	0·106	29
30	0·368	0·264	0·193	0·160	0·105	30
40	0·368	0·263	0·191	0·158	0·102	40
60	0·369	0·262	0·189	0·155	0·099	60
120	0·369	0·261	0·187	0·153	0·095	120
∞	0·370	0·260	0·185	0·151	0·092	∞

REFERENCES

FISHER, R. A. & YATES, F. (1948). *Statistical Tables*. London: Oliver and Boyd.

JOHNSON, PALMER O. (1949). *Statistical Methods in Research*. New York: Prentice Hall.

MERRINGTON, MAXINE & THOMPSON, CATHERINE M. (1943). Tables of the inverted beta (F) distribution. *Biometrika*, XXXIII, 73–88.

(*Manuscript received* 10 *February* 1955)

CORRESPONDENCE

THE EDITOR
British Journal of Psychology

Dear Sir,

In his recent paper,[1] Mr A. S. C. Ehrenberg says: 'Scott Blair (1949)[2] adopts a common point of view when he says that it is clear that three dimensions are required to describe space. However, this is not only not clear, but demonstrably false.'[3] Considering, as a two-dimensional example, my description of a symmetrical American city built on flat ground, with streets at right-angles, he says: 'one-dimensional descriptions could be made up, for example, as follows:

1 2 3 4 5 6 7 8
9 10 11 12 . . . .
17 18 19 . . . . .

With this latter map, one could find one's way around...'.

However, he has not, in this way succeeded in describing position in a plane by means of a single dimension. All he has done is to 'hide' one of the dimensions. To find my way to no. 18, for example, I require not only the number 18, but also the number 8, which tells me where to start looking for any number larger than 8. Of course, if the whole city consists of a convenient square of 8 blocks, this may be obvious; but, 'to describe space', one should use a method which is not limited, in this way, to finite distances of arbitrary magnitude.

In American hotels, all the rooms on, say, the 20th floor are often numbered upwards from 2000 (2001, 2002,..., etc.), even though there may be considerably less than 100 bedrooms per floor. Then no room is numbered 2099 and the first room on the 21st floor will be 2100 or, better, 2101. This is another convenient version of Mr Ehrenberg's system, or something rather like it; but, of course, each of these numbers is really complex. In the case of, say, 2143, the 21 represents directly the vertical dimension and the 43, a mixture of the two horizontal dimensions. Even on the 21st floor, we cannot tell where the room 43 will be without some other 'hidden' number which, like the '8' in Mr Ehrenberg's example, tells us how many rooms there are along each corridor, or gives us some similar architectural information. Thus no. 18 in Mr Ehrenberg's city should really be number 18:(8) and my room in the hotel 21:43:(45). The figures in brackets would, of course, require further explanation. In the latter case, (45) might tell me that my room was the third from the far end of the corridor.

The point is that we still have our two magnitudes to describe position in a plane and three to locate a point involving also altitude. (I do not, of course, imply that highly accurate measurements made over exceptional distances may not require the use of geometries more complex than that of Euclid.)

---

[1] Ehrenberg, A. S. C. (1955). Measurements and Mathematics in Psychology. *Brit. J. Psychol.* XLVI, 20–9.

[2] Scott Blair, G. W. (1949). *Measurements of Mind and Matter.* London: Dobson.

[3] Mr Ehrenberg was kind enough to make this criticism to me some time ago in a private conversation, but I could not immediately frame an adequate reply and I was most unfortunately not able to be present when he read his paper in Nottingham.

I willingly admit that my American city analogy, like all analogies, is imperfect. First, even the largest American cities have arbitrary boundaries and, as I have already said, we should not assume arbitrary boundaries in selecting fundamental methods for measuring space. Secondly, one might be tempted to think only in terms of discrete units—blocks, houses, etc., and to forget that, unless we are dealing with quantum physics, there should be no limit to the precision of our measurements. This point is covered by adding any necessary number of decimal places to our figures. Thus, instead of

$$\begin{array}{cccccc} 1,1 & 1,2 & 1,3 & . & . & . \\ 2,1 & 2,2 & 2,3 & . & . & . \\ . & . & . & . & . & . \end{array}$$

we can write

$$\begin{array}{ccccc} 1{\cdot}0,\ 1{\cdot}0 & 1{\cdot}0,\ 1{\cdot}1 & 1{\cdot}0,\ 1{\cdot}2 & . & . \\ 1{\cdot}1,\ 1{\cdot}0 & 1{\cdot}1,\ 1{\cdot}1 & 1{\cdot}1,\ 1{\cdot}2 & . & . \\ . & . & . & . & . \end{array}$$

If we could be content with discrete units, we might use a method somewhat like that of the 'arrondissements' in Paris.

$$\begin{array}{cccl} 7 & 8 & 9 & \\ 6 & 1 & 2 & \text{etc.} \\ 5 & 4 & 3 & \end{array}$$

We can make our units as small as we please by using enough numbers to cover any given area, but, whatever the size of the unit, the 'key' to such a method involves stipulating a 'quantum': the smallest area which can be located and, if we alter the size of the quantum, we alter the whole system.

We may be tempted to take a fairly large unit and to divide this into, say, nine parts, any number of times, allocating second, third and subsequent digits in a similar spiral fashion. But the selection of nine is itself as arbitrary as is the size of the original units and the method will not work unless some such number is chosen.

This is the nearest I can get towards describing location in a continuum plane by a single number. Perhaps Mr Ehrenberg can do better; but the matrix proposed in his paper certainly does not help.

Whether it is false to say that 'three dimensions are required to describe space' thus depends rather (as the late Prof. Joad would have said) on what you mean by 'describing space'.

I should add that with much of the rest of Mr Ehrenberg's argument I am in the heartiest agreement. Too rigid a theory of dimensions can have a crippling effect on the development of physics and, in some respects has already done so. Indeed my main purpose in writing my little book was to warn psychologists and others of this danger.

Yours faithfully,

G. W. Scott Blair

Physics Department
National Institute for Research in Dairying
University of Reading
17 *March* 1955

THE EDITOR,
British Journal of Psychology

Dear Sir,

In his stimulating letter, Dr G. W. Scott Blair suggests that the point 18 in my illustrative one-dimensional space-description might be called 18 (8). By this he seeks to demonstrate the existence of additional dimensions. But he is in fact developing rules and laws telling us how to get from one point to another, and in doing so has used only the values of the one dimensional scale, which therefore remains quite sufficient. Admittedly there are implicit in this one-dimensional system of measurement other, multi-dimensional ones. But then any one measuring system necessarily 'hides' all others that are possible. My own concern of course lay with the given measuring system, in its own right, overtly.

Dr Scott Blair's claim that 'we still have our two magnitudes to describe position in a plane and three to locate a point involving also altitude' would only be true—and then it would amount to a tautology—if the concepts of a certain kind of 'plane' (e.g. a horizontal one) and of 'altitude' were to be immediately explicit in the measuring system as defined: this dimension means altitude, those two define a horizontal plane. But with other measuring systems such concepts could make their appearance later, in the all-important development phase. This Dr Scott Blair has himself indirectly illustrated, and more generally, concepts like *inclined* planes, for example, would only emerge later even for the three-dimensional system that he has in mind.

The finer subdivision of a domain of measurement, and its possibly infinite extent—points which Dr Scott Blair has also touched upon—do not, I think, raise different issues for a one-dimensional space-description from those for a three-dimensional one. The theorem referred to in the paper, that all rational points are enumerable, seems to cover it all and since Cantor's original proof amounted to showing that rules for such enumeration exist, any attempt at their further elaboration would be an academic exercise which I myself would like to avoid.

Yours faithfully,
A. S. C. EHRENBERG

EDITORIAL NOTE

We have received a letter from Dr D. J. Finney, University of Aberdeen, commenting on the reference made to him by C. C. Anderson in his paper 'Some simple methods of testing for function fluctuation' (this *Journal*, XLVI, part 1, 1–12). Dr Finney points out that the 'memorandum' quoted was in fact a note prepared for the Director of the National Foundation for Educational Research in England and Wales (at his request) in comment on a confidential report submitted by Mr Anderson, and was intended solely as a basis for private discussion between the Director, Mr Anderson and himself. Dr Finney strongly deprecates any suggestion that his note represented a distinctive theory of function fluctuation and asks us to state that he would never have consented to a published discussion of a hurriedly prepared private document on a subject of which he lacked expert knowledge.

Mr B. S. Morris, Director of the National Foundation, has indicated that he is in entire agreement with Dr Finney's remarks.

Mr Anderson writes: 'I regret very much that my paper has given offence to Dr Finney. I was entirely unaware that his document was confidential and preparatory. If I have misrepresented his views I offer my sincere apologies.'

Mr Anderson has also written to correct the formula which appears on p. 8 of his paper. This should read:

$$\frac{\dfrac{B-C}{p}}{\dfrac{B-C}{p}+C}=\frac{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)}{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)+1}=\frac{I_2}{I_2+1}.$$

PUBLICATIONS RECENTLY RECEIVED

Probleme der ganzheitspsychologischen Wahrnehmungslehre. By WALTER EHRENSTEIN. Leipzig: Johann Ambrosius Barth. 1954. Pp. ix+342. Paper, DM. 19.50; bound, DM. 21.40.

This book, by one of the best known followers of the Gestalt school, is a new (third) edition, considerably changed from that which appeared in 1942. To experimental workers in the perceptual field it will be a store-house of brilliantly conceived experiments, interesting discussion, and somewhat biased theoretical speculation. It shares with its predecessor the same virtues and vices. Modern American literature is almost completely disregarded; theoretical discussions are twenty to thirty years out of date. The book certainly cannot be used, as the author claims, as a text-book; students would get completely erroneous views of modern theories in perception. Used rather as a compendium of interesting and often important phenomenological observations, this paper-backed volume is well worth its price. However, the reader would almost certainly be required to re-interpret the facts in terms of more modern and more consistent principles than those contributed by Ehrenstein.

The book may be of interest also to workers in the field of personality as the author has used his background in the perceptual field as a basis for a well-known typology. There has been considerable interest in recent years in linking up these two fields, and while the methodology used by Ehrenstein for the purpose of establishing his typology (cf. his book entitled *Grundlegung einer ganzheitspsychologischen Typenlehre*) is not adequate for the purpose, nevertheless students will find many hypotheses emerging from Ehrenstein's discussion which are worthy of being put to an experimental test.

H. J. EYSENCK

Physical and Psychical Research: an Analysis of Belief. By C. C. L. GREGORY and ANITA KOHSEN. Reigate: Omega Press. 1954. Pp. ix+213. 15*s.*

The main difficulty about psychical research is that its results indicate the occurrence of events which are contrary to the expectations based on the acceptance of an orthodox scientific system of belief. Two reactions are possible to this situation. One is to adhere to the belief system and to reject the alleged facts as impossibilities. The other is to question the foundations of the system of beliefs which has led one to make the judgement which leads to rejection.

Every increase in the strength of evidence for at least some paranormal phenomena is an additional reason for preferring the second alternative, and the strength of evidence has increased during the last thirty years to an extent not always realized. The authors of this book (a physicist and a psychologist) accept the necessity for the second of these alternatives, and they make a searching examination of the orthodox science belief system. Their position is summed up on the last page but one: 'Science as a discipline is magnificent and indispensable; as a belief system it is disastrous.' They contend that the implicit belief system of the Western World exerts a general defacilitation of images other than those related to material systems.

The authors envisage the possibility of a unified science which will include the facts of parapsychology and serve to bridge the gap between rival systems of belief. As a means towards this end, they suggest the necessity for the development of 'metascience' which would have as its subject-matter the informational relations between centres, individuals, and groups. The present book may be regarded as preliminary spade-work in preparation for these ends. It is to be recommended to those who are interested in the revaluation of their own belief systems.

R. H. THOULESS

Psychological Statistics. By QUINN McNEMAR. New York: Wiley; London: Chapman and Hall. Second edition. 1955. Pp. vi+408. 48*s.* net.

The first edition was reviewed in this *Journal* (1950, XLI, 87–8) by Patrick Slater. The author had not originally intended the book for use in introductory courses, and changes in the new edition are designed to make it more suitable for this purpose. Nevertheless, to the present reviewer Mr Slater's earlier criticism still appears to be valid: that the presentation is a little difficult for the mathematically unsophisticated.

In addition, some topics (e.g. Discriminant Function, pp. 210–11) are treated so concisely that one is left uncertain both of what the technique is designed to accomplish and of how the actual computation is carried out.

Throughout, the author's emphasis is on testing of hypotheses as the primary aim in the application of statistical inference, with the binomial distribution as the preferred starting-point of his exposition. The development of this theme has led the author to rewrite the chapters on analysis of variance, which now receive much fuller treatment than in any other text of comparable length and scope. Anticipated criticism of the short space (4 pp.) allotted to currently fashionable non-parametric methods is answered by the author's statement that in his view not enough is yet known about their respective merits.

Most users of statistics have their favourite text-books; while the present reviewer is unlikely to transfer his allegiance to McNemar, he has already found the book indispensable, and the Appendix contains a particularly handy selection of Tables.

BORIS SEMEONOFF

Transformation Analysis of Factorial Data. By YRJÖ AHMAVAARA. *Annals of the Finnish Academy of Sciences*, Series B, Volume LXXXVIII, 2. Helsinki. 1954. Pp. 150. No price given.

This monograph is in three parts. In Part I the author presents his 'transformation method', of comparing different factorial studies carried out with different populations. In Part II he applies this method to a comparison of several of Thurstone's studies of mental ability. Finally, in Part III, he outlines a technique by which the residual matrix in a factor study may be analysed to detect further 'latent' common factors.

The importance of the problem dealt with in Part I cannot be denied. The author tackles it by drawing a parallel between Pearson's selection theory and Thurstone's multiple-factor theory. Then by postulating 'a theory of factorial invariance', and using an argument similar to that used by Pearson he constructs a 'comparison matrix' by which the results of the two studies can be compared. More precisely suppose there are two $n \times r$ matrices, X and Y, of factor loadings one for each of two samples, where n refers to tests and r to factors. Then if the matrix X be taken as the invariant link between the factor scores for the two samples—corresponding to the invariance of the regression plane assumed by Pearson—and the specific factors remain orthogonal to the common factors and to each other, the comparison matrix takes the form $(X'X)^{-1}X'Y$.

This form is familiar. It is the rotation matrix required to rotate the factor matrix X into maximum conformity, in the least square sense, with the matrix Y. It can be obtained quite independently of Pearson's selection theory and of the assumptions made in it. (Indeed the author seems to do so on p. 51). The recognition of this fact is important and helps one to evaluate the author's rather far-fetched, if not misleading, statement (pp. 36–7) that 'the comparison matrix of given factors is independent of the population. This must even be the case, for what is concerned is simply that the result of comparison does not depend on the direction in which the comparison is made.' This statement is true only while the matrix X is retained as the invariant link between the two studies. This is equivalent in our terminology to saying that the matrix X has been chosen for rotation. Here an analogy may be drawn with regression, for just as there are two regression lines so there are two comparison matrices, and if comparison is made in the opposite direction the new comparison matrix is $(Y'Y)^{-1}Y'X$. The author no doubt realizes this but he certainly does not make it clear. Furthermore, the analogy with regression leads one to speculate that what is required, if this approach is adopted, is something like the product of these two comparison matrices, just as the product of two regression coefficients gives the square of the correlation coefficient.

Ahmavaara, however, justifies his method by using it very effectively to test the invariance of some well-known factor results by Thurstone.

There follows a useful critical review of the simple structure concept, and of what changes are to be expected in a factor pattern with increase in age. The monograph ends with three short chapters describing how the author examines a residual matrix for latent common factors by the calculation and examination of coefficients which he calls 'residual spectra'. In this way he is able to demonstrate, again using Thurstone's data, that latent common factors that are revealed there result from the tendency of the primary factors to break up into what he terms their 'fine structure components'.

In the monograph the number of errors that occur where reference is made to earlier pages and paragraphs are too numerous to mention. One safe rule would be to read §3 for §2, and §4 for §3, almost everywhere one meets them.

All in all this monograph provides very stimulating reading and this is especially true of the non-mathematical sections. One is indebted to the author for bringing into the open at least one of the greatest limitations of the factor technique, that is the need of a satisfactory statistical test for comparing results obtained from different experiments.

A. E. MAXWELL

The Metaphysics of Logical Positivism. By GUSTAV BERGMANN. Longmans, Green and Co. 1954. Pp. x+341. 42*s.*

Prof. Gustav Bergmann, one of the original members of the Vienna Circle, is known among psychologists for his application of the Carnapian technique of artificial languages to psychological theory. The emphasis of this work, is however, primarily philosophical, and most of the eighteen essays collected here have been previously published in philosophical journals.

Bergmann gives a good account as well as critique of logical positivism, which explains the 'metaphysics' in the title. He points out that two distinct types of linguistic philosophy have now developed: (1) devoted to the construction of artificial languages, cultivated to-day largely in the United States; (2) the analysis of philosophical correct usage which flourishes among the therapeutic positivists in this country. He is wise enough not to regard philosophical problems as being merely verbal, but he does imply that the technique of the artificial language is the *Open Sesame* to such problems. He believes, for example, that an adequate solution of the mind-body problem exists in terms of such a logical reconstruction of experience.

His essay 'On the problem of relations in classical psychology', is of direct interest to psychologists. All the classical psychologists discussed, British writers such as Locke, Hartley, James Mill, Reid, and continentals like Wundt and Brentano, accept a common method of inquiry, introspective analysis. These writers either conceived relations as analysable into atomistic elements, or as the resultant of a creative psychological act. They however overlooked another possibility which forms the very core of the more recent Gestalt doctrine, namely the irreducibility and immediacy of relational experience. Bergmann's criticism of classical psychology is then not that it was unscientific, but that it did not adequately take account of relations as immediately given experiential elements.

Besprinkled as it is with logical formulae this book is at times far from easy reading. Nevertheless, it will repay study by those psychologists who wish to obtain some insight into the linguistic vagaries of some modern philosophers.

W. MAYS

Die Frustration in persönlichkeitspyschologischer Beleuchtung. By REINO PALAS. *Annals of the Finnish Academy of Sciences*, Series B, Volume LXXXVII, 2. Helsinki. 1954. Pp. 110. 400 Markka.

This is a general survey of the part played by the concept of 'frustration' in contemporary psychology. It is defined as 'a state of affairs in which a living being is prevented by some obstacle, external or internal, from reaching its goal'. The author goes on to consider causes of frustration, reactions to frustration, and individual differences in ways of coping with frustration. His material is drawn from Freud, from the various offshoots of Freud's work, from Lewin and his pupils, and from recent American contributions. There are also references to Scandinavian and German writings which are less accessible to ourselves.

The author's success in keeping abreast of recent developments, especially in the U.S.A., compares more than favourably with that of most continental psychologists. One must, however, admit that there are signs of his remoteness from the places where most of the work he describes is going on. To, say, an American psychologist in the thick of current controversies, he must seem occasionally to have missed the point. His monograph is likely to be valuable as a convenient means of putting German-speaking psychologists in the picture. But it will not have much that is new to tell English-speaking readers, and more satisfactory summaries of the field are readily available.

D. E. BERLYNE

Human Development. By JOHN P. ZUBEK and PATRICIA ANNE SOLBERG. New York: McGraw-Hill. 1954. Pp. vi+476. Price $6.00.

The aim of this book, explicitly stated in the preface, is to satisfy the 'need for a text that would cover the entire age span from conception to old age, devoting about the same amount of space to each period of life'. The pattern of the book, largely determined by such an aim, emerges very clearly. Each section, whether it deals with the senses, learning, emotion or attitudes, traces the behavioural changes which occur as the individual progresses from infancy, or earlier, through childhood, adolescence and adulthood to old age.

Approximately half of the book is physiological: the first eight chapters deal with genetic factors in behaviour, and neural, glandular, physical, motor and sensory development. For the rest, learning, intellectual development, emotional development, social development, changes of interest, attitudes and personality are given a chapter each.

The scope of the book is thus enormous, and inevitably the pint pot will not hold several gallons. Yet one would hesitate to call this a superficial book: rather, the authors have chosen to deal with each topic from one particular point of view, the developmental, and to limit themselves severely to this. There are, however, some strange omissions—a chapter on learning and symbolic processes with no reference to Hull, Mowrer or Miller; a section on intelligence and mental factors with no reference to Spearman or Burt.

The major achievement of the authors lies in bringing together, in one reasonably sized volume, a tremendous amount of material, a great deal of it the result of recent research, which previously was scattered abroad in innumerable journals and text-books. They have collated this material and presented it in a well-organized text-book, lucidly written and very well illustrated, which should prove extremely useful to students of developmental psychology. The bibliography, containing well over 750 references, is evidence of the wealth of material provided, though in a book of this kind one might argue for a separate bibliography for each chapter, rather than one covering the entire book.

ELIZABETH FRASER

A Dynamic Psychopathology of Childhood. By LAURETTA BENDER. Springfield, Illinois: Charles C. Thomas. 1954. Pp. xiii + 275. 54*s*.

This is the third of the Bellevue studies in Child Psychiatry. It is a collection of papers by a number of authors, Lauretta Bender and Paul Schilder being the chief contributors. The papers have in the main been previously published but are enlarged here by the addition of later follow-up material which is of considerable interest. In some cases records run from early childhood into adult life.

No systematic survey of psychopathology is attempted, but the topics dealt with are of importance in psychopathology. There is stress on the nature of the fantasy life in children, and with studies of imaginary companions, hallucinations in children and internalization of fantasy objects. Studies of children's reactions to the realities of sexual seduction and to the death of parents or siblings yield some of the more interesting information.

In general, the number of children in each study is small, less than twenty, and the case-history of each child is summarized. The reader may in the end feel that he knows far too little about far too many children, while the small numbers naturally reduce the value of the conclusions. One may wonder, for example, whether the comparatively salutary effects of pre-pubertal sexual relations found in this study is general; or again whether the breakdowns associated with failure in sex-identification would be so apparent in a population containing fewer schizophrenics. However, much of the case material is of a kind not easily to be met with and its publication is clearly of great use. Where it is given in some detail, as in the chapter on Impulsions, Compulsions and Obsessions, its value is a good deal enhanced.

ELIZABETH NORMAN

Psychosurgery and the Self. By MARY FRANCIS ROBINSON and WALTER FREEMAN. New York and London: Grune and Stratton. 1954. Pp. ix + 118. $3.00.

This is Dr Robinson's book. Dr Freeman contributes only a very short chapter on the techniques of prefrontal leucotomy. But he was one of the first to see that leucotomy is apt to bring about real, if elusive, changes in self awareness and to direct his psychologist to their study. And he was singularly lucky in his psychologist. To her technical qualifications, Dr Robinson adds qualities of perceptiveness and humanity as admirable as they are rare. She has had the good sense to realize that clinical testing, to have any real value, must be founded upon—indeed grow out of—intimate clinical study. As a result, her tests are based not—as is usual—upon some arbitrary consortium of faculties, traits or factorial dimensions but upon close and sympathetic study of individual cases. Her experience—and Dr Freeman's—led to the hypothesis that an important effect of leucotomy lies in reducing what may be crudely called 'self-continuity'. By this is meant that the patient no longer relates himself fully to his past or his future, lacks all subtle appreciation of himself in relation to others, and no longer depends for his self-respect upon the acceptance or approval of his fellows. He is, in short, a simplified preparation of *Homo sapiens*.

The tests described are mainly directed to assessing in more quantitative form these changes in the leucotomized patient's attitude to himself, his past and his future. They are based largely on the extent to which the patient talks about himself (and to a lesser extent on the ways in which he talks about himself) in the course of a contrived experimental interview. Control studies are reported with comparable psychiatric patients who had improved without operation. The results undoubtedly indicate some failure of self-critique and diminished interest in one's own past and future as characteristic long-term effects of operation. The results of an additional method, quaintly called a 'Sensibility Questionnaire', point in the same direction. In general, the test results convincingly bear out the clinical observations which inspired them. They do not, however, throw any new light on the mechanisms responsible for personality change.

It will not escape the psychiatrist that the bulk of this work was carried out with patients who had undergone so-called standard leucotomy—in which the extent of surgical damage is fairly considerable. These operations are now increasingly giving place to more limited types of leucotomy, and it is notable that in the few cases reported here in which less extensive operations were carried out the disturbance in 'self-continuity' was correspondingly diminished.

This little book deserves the sympathetic attention of psychologists—and not merely of those whose routine work takes them into the clinical field. Without venturing an opinion on the rationale of psychosurgery, or on the vexed question of whether its psychiatric gains outweigh the losses, this book may be recommended to all psychologists not yet so atrophied by professional involvements as to have lost all curiosity about the human mind and its obscure dependence on the brain. Had William James been alive to-day, it is not too fanciful to suppose that he might willingly have written a Foreword.

O. L. ZANGWILL

Emotions and Bodily Changes. By FLANDERS DUNBAR. New York: Columbia University Press; London: Geoffrey Cumberlege. Fourth edition. 1954. Pp. xxii + 1192. £6.

This well-known book has now reached its fourth edition, which incorporates supplementary material and additional bibliography. The changes all appear to be in the nature of addition, without subtraction or substitution. The original theoretical introduction is still retained, although much of it has now rather an antiquated flavour.

The important new material is chiefly in the Clinical sections, under the headings of Metabolism (including problems associated with the feeding of infants and children), Genito-urinary System, Special Sense Organs (under which heading hysterical sensory abnormalities are presented), and Bones, which includes the section on Dentistry and an account of how its application in that field emptied a Dental Conference.

The main features of the book are familiar. It may justly be described as a mine of information, although not without a certain implication that some digging, sorting, and refining is necessary.

In his theoretical approach the author manages to combine the organismic and the psychoanalytical without discomfort. It is arguable, however, that the system implicit in the clinical sections is tacitly dualist.

The clinical reports, abstracted to the number of some five thousand, did not yield any controlled clinical inquiries to sampling, but this was perhaps merely the reviewer's misfortune. At all events, this type of inquiry cannot be said to bulk large in the field. No definitive answer can therefore be expected to theoretical questions.

The value of the book, then, to the academic psychologist, is largely limited to a surreptitious glance at the section on his own symptoms. To the clinical psychologist it is a five-thousandfold repetition of his own type of experience. To the practising physician, and perhaps more particularly specialist, at whom, presumably, the book is principally directed, its value could be considerable. This is, no doubt, the justification for the Clinical presentation, system by system and organ by organ, which otherwise appears somewhat clumsy and repetitive. Symptoms can be looked up as in the usual types of medical reference book, theoretical sections skipped, and suspicions crystallized.

Whether this leads to an adequate therapeutic approach is open to some doubt. This cannot, however, reasonably be regarded as a fault of the book, but is, rather, an unfortunate consequence of the spectacular triumph of the analytical approach in medicine over the past fifty years, of the short-term dividends that thinking in terms of 'diseases' rather than 'disease' has paid.

J. A. L. NAUGHTON

Medical Treatment of Mental Disease. By D. J. McCarthy and K. M. Corrin, with sections by eight contributors. Philadelphia, Lippincott; London: Pitman. 1955. Pp. xx+653. 96*s.* net.

This book is to be welcomed, if only because the authors see mental disorders as a part of general medicine to be treated as medical problems. They see psychiatric cases as sick people to whom one should apply the fundamental rules of general medicine. For this emphasis they are to be lauded. While a number of differences in style and viewpoint of the various contributors can be detected, the orientation of the book is fundamentally in favour of the organic as opposed to the psychogenic basis of psychiatry. They assume an organic or toxic aetiology for the vast majority of mental diseases, using this for the study and treatment of the case. The concepts and treatment of mental disease are dealt with as matters of disordered chemical, physiological or pathological processes affecting brain function. Emotional factors are recognized as being important, but more often as the effects of disease rather than the causes. No attempt is made to integrate the dynamic and purely psychological aspects. The authors believe they are restricting themselves to fundamentals of proven scientific value and hold the view that only on an organic basis can psychiatry advance scientifically.

In many respects, the book is out of date: the detection of hidden infection in the body is stressed as one of the most important aspects of psychiatric diagnosis. This emphasis on focal infection will to many psychiatrists be a return to the early part of the century when ruthless eradication of 'foci' was the order of the day. The section on personality is disappointing, as is that on examination of the patient. An extreme mechanistic approach is overemphasized by the constant reference to the human body as a 'physical machine' which must be got into a 'good shape of working efficiency'. Modern physical methods of treatment such as ECT and insulin coma are dubbed under the global title of 'The Drastic Therapies'. Among these the prominence and favour given to transorbital lobotomy using an electrically induced convulsion as an anaesthetic would hardly be viewed with favour in this country.

In spite of these criticisms this book contains much sound common sense. The medicolegal aspects of American psychiatry are well and interestingly dealt with. Psychiatrists who read it will probably be less likely to overlook coexistent physical disease in their cases. The stress on pharmaceutical therapy of mental disorder reminds us that drugs can control symptoms while other methods are given time to take effect. The book rightly emphasizes the role of the family doctor in the treatment of mental disease; and it is he, rather than the psychiatric specialist, who may find the message in these pages of value.

J. A. HARRINGTON

The Psychological Variables in Human Cancer. Edited by J. A. Gengerelli and F. J. Kirkner. California University Press and Cambridge University Press. 1955. Pp. iv+135. 22*s.* 6*d.* net.

This book reports a symposium held in 1953 in California, on investigations which tried to evaluate the relationship, if any, between human personality variables and neoplasms. That such a symposium should have been held at all will surprise many. The editors recognize that this is a difficult subject prone to misinterpretation and error, and rightly stress that the data presented should be viewed in a critical light until further research has been carried out. The main discussion centres round observations alleged to show that differences in the host's resistance to the growth and spread of malignancy are correlated with patterns of behaviour. The hypothesis is that the rapidity of the disease process, as measured primarily by survival time, correlates significantly with anxiety, defensiveness, and inhibition, as measured by psychological tests, chiefly the MMPI. Predictions using MMPI profiles were shown to be 78% correct in assessing fast- and slow-growing neoplasms. The classification into slow- and fast-growing tumours leaves much to be desired, and there are obvious criticisms centred round the validity of the psychological tests employed. Even assuming the MMPI to be a valid test for the purpose, and the correctness of the experimental findings, there are obviously other ways in which the results can be interpreted. The experiments deal with so many uncontrolled variables that spurious correlations may readily arise; observations on relationships do not establish causality. Even if it is true that a state of anxiety coincides with an increased malignant growth activity and a state of serenity with its retardation, it remains undecided which factor came first. Many of these criticisms are contained in the discussions following the papers.

Among the papers quoted, that by Schefflin on the incidence of cancer in institutionalized psychotics is of interest. He found the incidence of cancer in schizophrenics about half that in the general population and occurring at a later age; an exception being the paranoid schizophrenic in whom cancer occurs about four times as frequently as in the general population. The implication that cancer growth may be influenced directly or indirectly by psychological factors should not be taken too seriously in the present stage of our knowledge. Theories to account for relative immunity are legion and there are good *a priori* reasons to suppose that the solution will not come from the field of psychosomatic medicine.

The book makes one thing clear: that we lack the proper tools and techniques to investigate such a problem adequately at present. Until we do, it is doubtful if further extensive research in this field is justified.

J. A. HARRINGTON

Psychomotor Aspects of Mental Disease. An Experimental Study. By H. E. King. Cambridge, Mass.: Harvard University Press; London: Geoffrey Cumberlege. 1954. Pp. xiv + 185. 28*s*.

This is an interesting experimental study of psychomotor ability in chronic mental patients. A battery of objective tests was used, including reaction time, speed of tapping, and finger dexterity, and some 300 patients and controls were tested in all. Very marked differences were found, and the author quite rightly feels that the defects in mentally disordered patients unearthed here are of fundamental importance, from both the theoretical and the practical point of view. His conclusions certainly fit in well with those of others who have used objective behaviour tests on neurotic and psychotic groups of patients.

It is doubtful if this study really is worthy of presentation in book form at the price of 28*s*. The text is padded, partly by lengthy discussion, partly by excessive documentation in the form of figures and diagrams. The author justifies this on the grounds that it makes comprehension of numerical material easier for the uninitiated. This seems a doubtful argument; anyone unfamiliar with means and standard deviations is unlikely to derive much benefit from diagrammatic representations either.

H. J. EYSENCK

Psychotherapy and Personality Change. By CARL R. ROGERS and ROSALIND F. DYMOND. Chicago: University of Chicago Press. 1954. Pp. ix + 447. 45*s*.

The publication of this book, which reports some results of the Chicago experiment on the effects of psychotherapy, has been awaited with impatience by many psychologists who had heard of this ambitious project, or who had seen preliminary communications. Its final appearance will provide psychologists and psychiatrists with considerable material for thought and for discussion. In essence, what the authors have done has been to select a small group of clients who came to the University of Chicago Counselling Centre, administer a variety of tests to the subjects before and after therapy, and then study the changes in test performance which had taken place. Main reliance was placed on the Thematic Apperception Test and on personality questionnaires, the latter being presented in the somewhat novel form called *Q*-sort, in which the subject, instead of answering 'Yes' or 'No' to a number of statements regarding his personality, has to sort these statements in order of applicability to himself. No objective behaviour tests were administered.

An attempt was made to procure control groups along two lines. One was to have a group of normal people who were also given the tests, but, of course, who received no therapy; the other method used was called the 'own-control' group, in which patients are tested first, then kept waiting for 60 days without therapy, are tested again, then receive therapy, and are tested once more. A last test is applied to all these groups after a follow-up period of 6–12 months.

The outcome of the experiment at first sight seems to be very positive. There are considerable changes for the better after therapy; no such changes are reported for the control group; and only very small changes are found in the 'own-control' group during the waiting period. These results, however, cannot be accepted at face value. In the first place, the measures taken, i.e. questionnaires and T.A.T., are highly subjective and presumably subject to suggestion and social pressure; they would be acceptable as part of a larger battery, including objective and physiological measures, but they certainly are not adequate by themselves to provide acceptable proof of the effects of psychotherapy. Secondly, the authors fail to provide a proper control against which the improvement of the therapy group could be measured. The treatment of this whole problem is disingenuous. A control group made up of normal people obviously will not do; there is no reason to expect from them any improvement from the normal status simply because of the passage of time. As regards the 'own-control' groups, this might at first sight seem to present an adequate control because the same patient's progress is assessed during a non-treatment period and during a treatment period. This would be true only, however, if the two periods were comparable in length. I have been unable to find any detailed statement regarding the length of treatment period, but the impression given is that this period was considerably longer than the 60 days of non-treatment. Consequently, we might expect the effect of spontaneous remission to be much greater during the lengthy treatment period than during the short non-treatment period, so that the effects of treatment and spontaneous remission would become confounded. This would appear to be the greatest weakness of this experiment and one which makes it impossible to accept the results at face value.

A third argument is likely to be put forward by psychiatrists more familiar with the more serious types of neurotic disorder. Most of the 'clients' taking part in the experiment appear to have been suffering from very minor forms of disorder, which in this country, at least, would not be dignified by the term 'neurosis', and which would not, by and large, receive or request any kind of psychotherapy. One cannot necessarily argue from effects produced in what appear to have been fairly normal people, largely under situational stress, to what might happen with seriously ill neurotic and psychotic patients. The reviewer finds it difficult to evaluate this third point as no detailed evidence is given by the authors of this book on this crucial point.

In discussing what appear to be the three main criticisms of the experimental design, no mention has been made of a large number of small criticisms, most of which are mentioned by the authors themselves, and which may have a very powerful cumulative effect. There is no room to discuss these in detail, but the reader will be aware that at each point experimental rigour has been sacrificed to the putative interests of the clients. The decision to do so was based on ethical grounds which it would be pointless to discuss.

Carl Rogers, in his introduction, provides a fitting comment on this work. Adapting a saying by Robert Hutchins he says: 'It isn't a good research in psychotherapy; it's just the best that there is.' The reviewer cannot disagree with the first part of this sentence, but the fact that he is forced to agree with the second part is an adequate comment on the work that has hitherto been done in this important area of psychological research.

H. J. EYSENCK

Group Therapy for Mothers of Disturbed Children. By HELEN E. DURKIN, PH.D. Oxford: Blackwell Scientific Publications Ltd. 1954. Pp. xiv+125. 25*s*.

This short book by a lay therapist, who is also a lecturer in psychology, reports about 15 years of experience of a psychoanalytically oriented method of therapy applied in about 100 groups of mothers meeting once a week. In the author's experience this therapy contributes a great deal to the improvement of relations within the family, thereby helping and safeguarding independent therapeutic work with children. Any serious effort towards these ends must claim attention from child guidance clinicians and family case workers.

The work reported is clearly that of a sensitive and experienced therapist, and the discussion and brief case illustrations contain many insights and suggestions which group therapists will find useful. Of particular interest are some suggested criteria for selecting members of these groups, the use made of transference relations within the groups and the discussion of the suitability of this form of therapy for 'character analysis'.

But taken as a whole the book falls far short of the claims made for it in Slavson's brief and somewhat pontifical Foreword. It makes only very limited contributions towards the development of a rationale for group psychotherapy, and the possible differences between the approach used with the mothers' groups and that in group therapy in general are not discussed at any length. We would like to understand much more about the effect on the therapeutic process of orienting the group work towards the mother–child relation, for the use of this fundamental relationship as a focus of interpretation, in transference and outside the group, may have important implications for therapeutic practice.

H. PHILLIPSON

Group Discussion. Edited by JOHN BURTON. London: Central Council for Health Education. Pp. 92. 3*s*. 6*d*.

The theme of this short symposium is that of the merits of education by discussion as against the passive nature of more formal didactic educational methods.

It opens with several articles on the organization and running of discussion groups in general, and follows these up with an account of varied applications of the group discussion technique. The advice in the opening articles is presumably addressed to the would-be group leader and ranges from the formal procedure involved in running group meetings to thumbnail sketches of types of group members with which the leader might have to deal (e.g. The wanderer, the silent, the shy, the monopolizing, etc.). The remainder of the symposium deals with the application of group discussion to the teaching of school children, youths, students, the general public and industrial workers, and closes with a discussion of group psychotherapy.

The majority of these articles, although interesting and informative, overemphasize the role of the leader in such a way that he is too often seen as an all-seeing, omnipotent eye functioning outwith the group. The point that the leader is at the same time a member of the group, subject to the same forces as other members, tends to be neglected. An article by Dr Polanie on the teaching of students by group discussion does deal more cogently with the function of the leader. He warns against the danger of the leader viewing the discussion as a group tutorial and stresses the leader's main aim as being—'to merge with the group and disappear'. Dr Anthony's discussion of group analysis contains much that is relevant to the functioning of any group of people but, unfortunately, this is the last article in the symposium and its relevance to other types of activity groups previously mentioned may not be too obvious.

Two articles merit particular mention for their originality and interest. Miss Kelbie describes an interesting technique of spontaneous group conversation used first with post-war refugees in Germany. Here the members of the group are given the general theme of childhood memories at a particular age and encouraged to make personal contributions. She points out the advantage that all can participate easily and that the happy memories shared are easily projected into the group to promote positive friendly group feelings. The dangers of negative feelings resulting are also discussed. The basis of this technique has much to link it with that of group analysis in a more superficial and direct form.

After the initial articles dealing with the form of organization and running of discussion groups and committees, etc., and detailing the different roles of leaders and members within these groups, it is somewhat of a relief to read Beatrice Pollard's interesting account of the Quaker Group: 'There are no rules, no standing orders, no rules of procedure, no motions, no commitments, no closure motions or any of the other apparatus with which the world of committees is familiar.' The Quaker group, in fact, appears to represent a halfway-house between the formal business meeting or committee type of group and the purely psychoanalytic group.

A. MCGHIE

Time Distortion in Hypnosis. By LINN F. COOPER and MILTON H. ERICKSON. Baltimore: Williams and Wilkins Co.; London: Ballière, Tindall and Cox. 1954. Pp. ix+191. 31*s.* 6*d.*

Part I (158 pp.), by Cooper, reports 'pilot experiments' (some previously published) with fourteen subjects, mostly students, on phenomena during 'time distortion' and 'time manipulation' in hypnosis: viz. injection of sound signals into hallucinatory activities; learning, motor and non-motor; review and practice of previously learned material; association; mathematical and creative mental activity. 'Time distortion' is defined as increase of seeming duration over clock reading, although the subject experiences the normal rate of activity; its most important implication is the possibility of *accelerated mental activity*, which would obviously be of tremendous practical significance.

Part II (28 pp.), by Erickson, describes therapeutic application in 6 patients.

This is at best an amateurish attempt at research. There is no reference to the experimental work on time perception, hypnosis, thinking, etc., only to the authors' own, 'communications', 'suggestions', etc. No objective evidence of accelerated mental activity is adduced, only anecdotal data, e.g. counting 821 jelly beans from a shoe box in a seeming duration of 10 minutes, actual time 10 seconds. The clinician's defence is the false dichotomy between scientific and human interests. The fundamental fallacy is that time is regarded as a *thing* to be measured instead of as itself a *measure*, 'experiential' time being but an individual's estimate of 'physical' time. Since its nature is not yet understood, to use hypnosis as instrument instead of object of research is premature. What these workers have done is merely to have tested the effectiveness of their hypnotic suggestion, contributing nothing towards understanding the status of hypnotic experiences.

J. E. MACDONALD

The Hands of Children: an Introduction to Psycho-chirology. By JULIUS SPIER. Second Edition, with an Introduction by C. G. JUNG. Translated by VICTOR GROVE. Illustrated with 86 drawings by F. W. W. SIMON and 69 Plates. London: Routledge and Kegan Paul Ltd. 1955. Pp. xvi + 199. 28*s.* net.

The first edition of this work was published in 1944. The new edition contains a second appendix by Herta Levi on the hands of the mentally diseased, with 46 plates to illustrate it, in addition to the first appendix on specimen analyses on the hands of children, which is illustrated by 23 plates.

The body of the book consists of eleven chapters setting forth the theoretical foundations of psycho-chirology and its application to the hands of children. To the empirical and experimental psychologist a book of this kind offers many problems and little help. It has developed out of traditional chirology and uses its language. Like graphology, it has adopted Jung's psychology as its theoretical background, and claims to be the intuitively established basis of a scientific approach to the psychology of personality. Exactly how much of it would survive strictly scientific and experimental tests, if they could be applied, remains problematic, though it contains numerous significant observations. Herta Levi shrewdly points out in the preface to the second edition that, as a scientific method, chirology is still in its infancy and requires a good deal of body-building nourishment in the shape of factual data and statistics, but adds that very few of us have the great intuitive powers of a Julius Spier, and therefore have to rely more on scientific methods and exchange of information!

R. W. PICKFORD

Low Intelligence and Delinquency. By MARY WOODWARD. London: Institute for the Study and Treatment of Delinquency. 1955. Pp. 24. 1*s.* 6*d.*

This pamphlet reviews a 40-year old controversy on the relation between low intelligence and delinquency. There is also a brief report of an intensive study of a small group of delinquents. The conclusion is that 'low intelligence cannot be regarded as an important causal factor in delinquency': whether intelligence has any relation at all to delinquency is not settled, owing to the difficulty of controlling the cultural factors which accompany delinquency and depress test scores. This conclusion is probably not surprising; but the pamphlet performs a valuable service in its concise review of the literature, summarizing 66 references in 16 pages. On this topic, successive investigators obtained conflicting results, a state of affairs not unknown to the social sciences, which tends to discredit research. The writer tries to resolve inconsistencies and present findings from the accumulation of results.

J. D. NISBET

Culture and Human Fertility. By FRANK LORIMER. UNESCO. 1954. Pp. 514. 25*s.*

The book divides into two sections, the first being a theoretical inspection of the relation of cultural aspects and human fertility, while the second presents a number of special studies in this field contributed by other collaborators. The reports in the latter section illustrate clearly the pitfalls which beset the unwary purveyor of vital statistics in areas where not only are reliable figures unavailable but where the population may view with darkest suspicion any attempt at enumeration. Prof. Mortara points out this absence of reliable statistics for Brazil, and the other studies reveal the perplexities arising for workers in preliterate communities in this respect. Further, the question of fertility has great emotional significance to many peoples, and this may lead to distortion or suppression—if not complete misrepresentation—of facts. In short, local conditions and attitudes may combine to produce data of low validity. Thus the need for caution in interpreting the results of such studies and the need for other carefully regulated investigations to supplement them must be emphasized.

In the theoretical section Prof. Lorimer draws on a wide range of material. But, as in much comparative work at the present stage, his exposition tends to show up the gaps and uncertainties in our knowledge. Believing in the pervasiveness of sociological factors, the reviewer agrees with his main thesis that these factors may exercise a profound influence on human fertility. The studies quoted appear to indicate this assumption. However, at present it seems impossible to do more than to formulate fairly general hypotheses, as he does.

The book, then, serves as a valuable exploratory work. The evidence is at times a little thin and the statistical material somewhat dubious. It must be considered as blazing a trail, not as a definitive assertion of the complex inter-relationships between culture and human fertility.

MARY BIRD

Sex and Society. By KENNETH WALKER and PETER FLETCHER. Harmondsworth: Penguin Books. 1955. Pp. 251. 2*s.* 6*d.*

This book is concerned largely with the question, 'What is wrong with sex in the modern world?' The authors take the view that much is wrong, and that sexual disorders are symptoms or aspects of unsatisfactory conduct of life in general, which in turn is due largely to the inhibiting and distorting effect of attitudes widely engendered in a highly organized 'mass society' upon individual spontaneity and values. This is regarded very seriously indeed—as 'a matter of life or death' for our civilization. While one may have sympathy for the general thesis (which is by no means new) one may well doubt whether the situation is as bad as the authors make out, and whether the outcome is likely to be as disastrous.

In the course of the book, many sensible things are said, as in the chapter 'Sexual Morality before Marriage', and some less sensible things, as in the chapter 'Introduction to Neurosis'. There is a tendency for 'typical' cases to be referred to without any evidence as to how typical they are, and for statements to be made without adequate basis being indicated; for example, 'It is not surprising that emotional maturity is a rare phenomenon in the modern world.'

D. GRAHAM

Working-class Anti-Semite. By JAMES H. ROBB. London: Tavistock Publications Ltd. 1954. Pp. xiv + 239. 15*s.* net.

In this study the author interviewed a near-random sample of 103 Bethnal Green working men and rated them on anti-Semitism and a number of personality variables, such as degree of social isolation. He obtained information on experience of unemployment, presence of Jewish workmates and so on; and attempted to relate the anti-Semitic personality to the background of the subculture of the district.

The results show that only one man in eight of the population is completely tolerant, but over half the remainder have to be openly invited to express their anti-Semitism before they will do so. The correlations between anti-Semitism and the personality variables rated were modest (0·33 to 0·57) but in the expected direction; no significant relations were found between anti-Semitism and any of the environmental factors investigated. In general the findings, such as they are, do not dispute any of the current views in this field.

The most interesting and potentially useful technique to come from this work, is the author's method of rating anti-Semitism according to the amount of provocation required to elicit a prejudiced remark. This variant of the focused interview would seem to have many uses, not in this field alone. It is thus all the more unfortunate that no attempt was made to validate the *a priori* classification used against, say, questionnaires or systematic case studies. The problem of the validity of making personality assessments from a 25-minute doorstep interview is nowhere tackled; and most of the correlations presented are suspect on the grounds of non-independence of the raters.

One further inadequacy must be noted; the author's attempt to relate the psychological and the sociological material. This is done by means of the concept of 'typical personality'. For the social psychologist, whose interest is in measurement in the social sciences, 'typical', if it means anything, should mean the mode of a modal distribution. Its use should be restricted to cases where this distribution has been first established; otherwise explanation in these terms becomes too easy. 'So easy', as Cattell said in another connexion, 'as to be worthless'.

In spite of these criticisms one must welcome this book as the first attempt to study anti-Semitism in this country on anything other than the questionnaire level. In a field where we have happily become used to the sort of on-going team researches of the California and Maudsley groups, it is inevitable that the contribution of one man working for two years should be considered as little more than a brave try. The book does serve, however, to emphasize the need for co-operative research and highlights the difficulties of a psychologist tackling this task alone.

DEREK PUGH

Christian Life and the Unconscious. By DR ERNEST WHITE. London: Hodder and Stoughton. 1955. Pp. 190. 10*s.* 6*d.*

In assessing the value of any book two things need to be kept in mind: (1) What it contains—not what it doesn't; (2) For whom it is written. Dr White's book is an attempt at showing how the Christian faith can influence the levels of the unconscious through the operations of the Holy Spirit. He distinguishes between Conversion and Re-Birth which works 'secretly' in the true Convert's mind. On the whole the book is orthodox on the religious side, and shows a distinctly eclectic analytical approach on the psychological. But the author has in mind Christians who are puzzled by the apparent dichotomy between religion and modern psychology. Readers who are biased against the former will find it difficult to appreciate the fundamental premises of the author, and those prejudiced against the latter are likely to reject the necessity for defining the mind in the way he does. This does not mean that the work is without merit, because it has considerable value for those who are willing to read it sympathetically. At times one gets the idea that Dr White's own religious beliefs intrude upon his work as a medical psychologist, but it is hardly likely that he is guilty of proselytizing, for he must know that no lasting good comes from the imposition of ideas on sick minds. The whole aim and purpose of the book is positive in the sense that it shows how religion can be aided by psychology, and how the human personality can be enriched through faith and religious experience.

W. E. SARGENT

A Dictionary of Pastoral Psychology. By VERGILIUS FERM. New York: The Philosophical Library. 1955. Pp. xi+336. $6.00.

The comprehensiveness of this book and its breadth of description could earn for it the perhaps more impressive title of encyclopaedia. In addition to the large number of psychological terms adequately defined, many subjects within the field of pastoral psychology are expansively treated, e.g. Pastoral counselling occupies 41 pages—a small book on its own. The volume is fairly well up to date, and could be extremely helpful to ministers in search of reliable principles on which to work when confronted with human problems, or in understanding new concepts relative to a pastor's life. The many cross references lend clarity and there are pertinent suggestions for further reading. A number of authorities, all specialists in their own fields, have contributed specific articles. Being an American, the author naturally quotes many of his own countrymen; one wishes, however, that a little more room had been given to writers in the British field of Pastoral Psychology.

W. E. SARGENT

Answer to Job. By C. G. JUNG. Translated by R. F. C. HULL. London: Routledge and Kegan Paul. 1954. Pp. xviii+194. 12*s.* 6*d.* net.

This book makes difficult, but rewarding, reading to the reader who has some acquaintance with Jungian thought. To the reader who is not familiar with that discipline, it appears strange and puzzling. To any reader, it would be much more rewarding if the translation were better: the language is uneven, ranging from the obscure to the colloquial, and it often jars unnecessarily.

The theme centres round the existence of evil, and the development of God's nature and man's, both of whom need awareness of their own darker side in order to achieve wholeness. Job, by his greater awareness and insight, offers a challenge to God, and the book thereafter, with great erudition and width of allusion, traces the development of self-awareness and of the archetypes in God—good and evil, the masculine ideal of perfection and the feminine ideal of wholeness—through the prophecies, the Incarnation, the Apocalypse and finally the doctrine of the Assumption of the Blessed Virgin Mary. The psychological aspect of the latter is interesting and important and dealt with here for the first time.

Atomic Warfare, and the way in which man's intellectual development has outstripped his spiritual growth, are shown as evidence of the urgent need to consider these matters.

Jung himself points out that the psychology of religion has two aspects, the psychology of religious persons and the psychology of religious 'contents'. He has himself, in this book, made a rare and original contribution to the latter.

A. M. SILVER

Olive Schreiner: a Study in Latent Meanings. By MARION V. FRIEDMANN. Johannesburg: Witwatersrand University Press. 1954. Pp. vii+67. 8*s.* 6*d.*

This book is an interesting attempt to apply psychoanalytic hypotheses to the interrelationships of neurosis and creative art. The personality of Olive Schreiner, as partially reflected in the characters and action of her novels, and as shown in her letters and life history, is dissected with some skill. These two sources of evidence are functionally correlated in the chapter on 'Latent Meanings'.

The study suffers from the usual disadvantages inherent in its approach. Any unusual work—whether 'art' or not—is generally the product of an unusual personality—statistically 'abnormal'. It is in the application of qualitative criteria of abnormality, therefore, that the approach adopted by Miss Friedmann may be valid. The neurotic is, compared to the artist, relatively helpless in the face of his difficulties, and resorts to inadequate forms of behaviour, which the clinician has little hesitation in labelling as pathological. There is some doubt in the mind of this reviewer whether this word can, without undue simplification, profitably be applied to the creative activity of the successful artist.

In the present instance, Miss Friedmann has thrown considerable light on the reflection of Olive Schreiner's own life in her fiction, though little is made of the political works and influence of the woman who has been called 'the first suffragette'. Reaction against oppression, even though unusual in the South Africa of 1900, cannot be regarded simply as maladaptive behaviour.

This book does not nor can it be expected to, answer the question 'Why was Olive Schreiner a great artist?', but it illuminates what she wrote, and contributes to our knowledge of what her writing meant to her. This year is the centenary of her birth and this study, though slight, is a not entirely unworthy monument to its subject.

S. G. LEE

The Function and Training of Mental Nurses. By A. N. OPPENHEIM. London: Chapman and Hall. 1955. Pp. 85. 12*s*. 6*d*. net.

Twenty years ago the duties of mental nurses were restricted to the custodial care of the mentally sick. With the introduction of complicated physical treatments, the mental nurse has had to fulfil duties more like his or her general trained counterpart. More recently, mental nurses in some Units have had to fulfil the duties of 'social therapists', in which the traditional skills of the hospital nurse are no longer relevant. It is little wonder then, that some confusion prevails as to the functions of the mental nurse. This in turn has an urgent bearing on training.

Mr Oppenheim's occupational study of mental nursing at the Bethlem Royal and Maudsley Hospitals is therefore timely, and it is not surprising that it has received considerable publicity in the Nursing periodicals. His study has been thorough and painstaking, and could well be taken as a model for similar studies in other hospitals.

According to an Appendix on Formal Training at the end of the book, the 16 lectures on Psychology included in the nurse's three-year course are given by the Sister Tutor and by 'a doctor'. One wonders why Clinical Psychologists are not asked to give any of these lectures.

RALPH HETHERINGTON

SIR GODFREY THOMSON

OBITUARY NOTICE

SIR GODFREY THOMSON

1881–1955

> Over Ottercap Hill they cam in
> And so down by Rotherley Crag
> Upon Greenleighton they lighted down.

Sir Godfrey's heart, it may be, rested in the hills and moors of Northumberland. One of his earliest papers was on the influence of Northumberland's geography upon its history, in which he traced the paths of marauding Scots, and of the far-spreading Romans, upon his northern county. The Scots might taunt a Percy under the walls of Newcastle: but dead would be the Douglas at Otterburn. The stirrings of the border, it seems, found deep echoes in Sir Godfrey's life, his work, his personality. He loved regional things. There was a delight and pride, perhaps, in his belief that the hill stock of Northumberland was of superior quality, descended not from a peasantry, but from the bold 'reivers and raiders of the marches', the men of Rotherley, Carter Bar, and Greenleighton, 'who pitted their wits against the equally alert Scots across the border'. Little is ever heard of the raids of Northmen into the wilds of Roxburgh or beyond; but it may have been with some such thoughts that Godfrey Thomson left Newcastle for Moray House thirty years ago—taking with him, however, chivalry and keen intelligence, scholarliness and wisdom, to live and commune with the Scots as one of them.

There are others, I am sure, who are better qualified than I am to appraise the man and his work. But I knew him in some sort of a personal way for over thirty years, and heard about him before that when I was a boy. As Professor Thomson he was my first mentor, and I was perhaps the first of his students of Newcastle days to pursue some of his interests. There are other bonds connecting us, some academic, others inconsequential but warm for the soul, and these matters perhaps give me the privilege to write about him. The young Godfrey went to Rutherford College in Newcastle, and then studied mathematics and physics at Armstrong College. There was a common thread, to which he himself used to draw attention, between his early invention of a compound pendulum and his childhood fascination with clocks and watches, and between these matters, for which it was always possible to develop elegant mathematical expressions, and his later interest in sampling theories. The beating out of sine curves is perhaps not very different from the complex throbbing of a human brain. A single strand of motivation, therefore, seems to run through all Sir Godfrey's work, from his earliest paper on a pendulum, to his latest exquisite tailpieces on *The Geometry of Mental Measurement*. What the motive stemmed from no doubt lay deep, if not inquiet, in his heart. But what it fashioned is wide open for anyone to see. For it is given to very few to combine constructive judgement and brilliant exposition as completely as did Sir Godfrey. His *Factorial Analysis of Human Ability* is still by far the most readable account of factor analysis available to us; and to expound these intricate matters so cogently is, as Thackeray said of Macaulay, to read twenty books to write a single sentence, or to travel a hundred miles for a single line about a country's beauty. Sir Godfrey was always supremely lucid. His expositions are crystal clear. But to this he added a fine eagerness for truth, and a sharp revolt against arbitrary assumptions. His Northumbrian forthrightness was never better put to service than in his famous stand against all arbitrary assumptions in Spearman's Theory of Two Factors. But for every critical probe there was always added an eagerness to acknowledge the merits of any theory, and a punctiliousness in giving credit where it was due. It was characteristic of the time, only thirty or so years ago, that these men could not converse freely with one another. The issues dividing Spearman and Thomson loomed too large, like monstrous ogres. Yet each, as I know, had enormous admiration for the other.

A full or balanced appraisal of Sir Godfrey's work is scarcely possible at this juncture. His early interests in physics led imperceptibly to empirical and theoretical work on the psychophysical methods, and thence to his notable chapters in *The Essentials of Mental Measurement*, for which he collaborated with William Brown. The *Essentials* is now of scholarly rather than immediate scientific interest, though the chapters on the sampling theory of ability represent Sir Godfrey at his most cogent and creative. His main point, made first by him I believe, was that innumerable factorial analyses are possible for data, and that any choice between them must be, as he said, on psychological grounds. Factors, he argued, are 'not necessarily anything real, they are only pointer-readings, mathematical symbols'. He agreed with Sir Cyril Burt in this matter. It is hard, indeed, to believe that anyone ever thought otherwise, but sophistication in this matter seems to have grown slowly, although prior to and independent of the increasing knowledge that is now available about the logic and philosophy of science. Factors, however, may be directly empirically rooted—they are then usually called 'descriptive' only; in this sense they are strictly abstract models of something real. Hypothetical factors of this kind are most directly employed in the structuring of samples, as in Q-method, and they are then indeed merely 'pointer-readings, mathematical symbols'. But there is a third, and perhaps a different, possibility for factors, which may have been Sir Godfrey's major concern. This stemmed from his belief that the mind is not divided into 'unitary' factors, but is a 'rich, comparatively undifferentiated complex of innumerable influences; on the physiological side an intricate network of possibilities of intercommunication'. The 'unitary' factors which he rejected, if they meant anything, must be of the theoretical kind, to which I have just referred. What, then, remained for Sir Godfrey? He found the answer in the work of Sir Cyril Burt, doyen of English factorists. In a review of Burt's *The Factors of the Mind* (*Brit. J. Educ. Psychol.*, 1941, XI, 46) Sir Godfrey expresses deep interest in Burt's comparison of factor analysis and the theory of groups used by theoretical physicists in quantum theory. The concern in the two cases is with the structure or patterns contained in action, with relationships or structure 'which exist between unknown and unknowable components'. Factors here merely specify systems of relations, without theoretical content of the kind we can build into Q-samples, or such as might be represented by lines of latitude and longitude on a map. What seems to have interested Thomson was the feeling, which he shared with Burt, that the final goal of factor analysis must be just such ultimate explanations, in terms of components which are perhaps unknown and unknowable. Sir Godfrey, however, liked the physiological model of networks of intercommunication in a brain; his sampling theory of intelligence represents this very well, and it may remain for generations as an acceptable and ultimate explanation of very complex matters indeed.

Sir Godfrey's knowledge of factor analysis was both minute and prodigious. But there is also a more general side of his work, which is also statistical, which was no less thoroughly pursued. It shows best in his Chairmanship of the Scottish Mental Survey Committee, with its studies on intelligence and its ramifications into fertility, social class, occupational status, twinning, family position, test sophistication and the like. He brought to these matters wisdom and caution. The complexities of the community and family, clearly, matched those of the brain and its networks of intercommunication. Statistical sophistication was essential to both. In these two preoccupations, then, one a sampling theory of intelligence, physiologically conceived, and the other the statistical analysis of complex sociological matters, we can see both the great strength and the great weakness of Sir Godfrey's work. His strength was to bring fine scholarship and mathematical rigour to bear upon these statistical matters. His influence on Ferguson, Fraser Roberts, Emmett, Lawley, Ledermann and many others is a rich testimony to this strength, and something about which the University at Edinburgh cannot but be proud. The weakness, so it seems to me, was that in between the intricacies of the brain on the one hand, and of the sociological matters on the other, human behaviour as such seemed somehow to be overlooked. But science is long, and life short, and everyone has a choice to make amongst many possibilities. Sir Godfrey, like Florence Nightingale long before him, was a 'passionate statistician'. He clearly held, with Quetelet, that 'we can judge of the perfection to which a science has come by the facility, more or less great, with which it may be approached by calculation'.

Of Sir Godfrey's great gifts as a teacher others have spoken. His written word was no less brilliantly espoused. 'No more than a bird can fly with one wing, or a man breathe only by inhaling' is a typical allusion from his *Instinct, Intelligence, and Character*, on some fallacies in reasoning. And as a chairman of meetings he was surely incomparable. When, as a student, I bumped into him on a stairway at Armstrong College—he was with his own sage mentor, Professor Mark Wright—his bearing was somewhat military, his personality a little, shall we say, superior or haughty. The perception of him was no doubt jaundiced, as any student's is apt to be. But I think that I discerned, over the years, a profound mellowing and growth of personality in the man. I saw him last seven years ago, and I felt then, as I do now, that one could say about him what was written about Washington Irving: 'in his family gentle, generous, good-humoured, self-denying, affectionate; in society an example of complete gentlemanhood; unspoiled by prosperity; never obsequious to the great; eager to acknowledge every contemporary's merit; always kind and affable to members of his calling; a charming master of language....' Such was the Godfrey Thomson I knew from a distance, for a long time.

W. STEPHENSON

A COMPARISON OF THE PERSONALITY PROFILE (16 P.F.) OF EMINENT RESEARCHERS WITH THAT OF EMINENT TEACHERS AND ADMINISTRATORS, AND OF THE GENERAL POPULATION

By R. B. CATTELL and J. E. DREVDAHL
University of Illinois and University of Minnesota

I. *The setting of the problem* (pp. 248–249). II. *Research plan and procedure* (pp. 249–251). III. *Differences from the general adult population* (pp. 251–256). IV. *The differences of researchers from administrators and teachers* (pp. 256–258). V. *Summary and discussion* (pp. 258–260). *References* (p. 261).

I. THE SETTING OF THE PROBLEM

Research on research is a recent conception, but it has come none too soon. As the investigations of Wolfe (1952) indicate, the well-springs of professional talent never flow too freely, and of all professional groups that of creative researchers is probably the most uncertain in its supplies of talent. Nevertheless, because of the importance of this small band to continued social prosperity and scientific advance, its occupational selection and conditions of working creatively need to be studied, despite their complexity, as soon as understanding of personality measurement and the mental processes of creativity (Kretschmer, 1931; Spearman, 1933; Guilford, Wilson & Christensen 1952), make possible an improvement upon present practices.

Probably more can be contributed at this moment through selection than merely by multiplying hours spent on research. Swelling research endowments without increasing our power to select researchers may merely drown the light of understanding. And unless good researchers can be selected younger we have the sad alternatives (*a*) of waiting till a man has proved himself by his works, which permits him to begin life at forty, (*b*) of pouring endowments on anyone whose birth certificate proves him to be young, a despairing and wasteful procedure, but one which any research council is at times tempted to follow, or (*c*) of judging research by the design submitted, rather than by the person, a practice which makes the research world safe for all, including all rule-following mediocrity.

Three well-known attacks upon the problem, which have recently rallied interest, are those of Guilford *et al.* (1952) upon the nature of creativity, and of Roe (1953) and Van Zelst & Kerr (1954), upon the personality of the researcher. This article does not purpose to survey and integrate, but we may point out that our objective is substantially the same as that of Roe, though our results cannot be aligned, because Roe had to depend at the time of planning her research on the Rorschach and the T.A.T., whereas our measurement is directed to personality dimensions more recently factor-analytically investigated (Cattell, 1950*c*) and confirmed (Cattell, Day & Meeland, 1953; Cattell, 1955*a*) in the Sixteen Personality Factor Questionnaire (Cattell, 1955*b*).

While this measurement of primary personality dimensions has the advantage of greater reliability (including lesser subjectivity of scoring) than the Rorschach and T.A.T., and also of higher validity in the sense of accounting for more of the variance of more demonstrably independent directions of personality variation, it has the disadvantage of all questionnaire techniques: possible motivational distortion and possible undue demands on literacy. Distortion has been avoided, as far as possible, in the 16 P.F. by indirection of questioning, but with persons eager for a particular job, some distortion undoubtedly remains for correction, and low educational level can also produce random error. It may be claimed, however, that the present investigation of teachers, administrators and researchers presents the most ideal situation for the use of the questionnaire, in that these subjects are educated people, appreciative of the scientific need for candour, secure in their positions, and assured of anonymity in their results.

Vocational selection and guidance depend in principle upon either or both of two statistical-conceptual systems. In one the personality-ability *profile* is determined for the typical member, by such devices as used by R. L. Thorndike (Thorndike & Norris, 1952), Cronbach & Gleser (1953) or one of the present writers (Cattell, 1949) in the formally identical problem of classifying culture patterns (Cattell, 1950*b*). Any individual propositus then has the similarity of his personal profile to this successful (or at least surviving) type determined by the pattern similarity coefficient, r_p, simple or weighted (Cattell, 1949; Cronbach & Gleser, 1953). In the other, a criterion of success within the occupational group is set up, and regressions, linear or curvilinear, of the various personality factors (including abilities) upon this criterion are worked out. Enlightened use of the values from either requires it to have the signature of the particular social supply-and-demand conditions obtaining at the time.

The present investigation was planned with the first of these objectives—to determine the personality profile of the typical researcher in relation to the average adult and the profile types of other occupations, especially of those for the related callings of scientific teachers and administrators, among which distinction most needs to be made in guiding students to careers. It seems to us that this delineation of the occupation *per se*, relative to others, should precede, logically and statistically, the investigation of the variance *within* the occupation. Besides, the preparation of criteria of success, involving, if it is to be effective, a good deal of time on the part of high-level professional people, was too expensive for the present, unendowed research.

II. RESEARCH PLAN AND PROCEDURE

The testing instrument, having been frequently described in other investigations (Cattell, 1950*a*, *b*, 1955*a*; Cattell *et al.* 1953; Drevdahl, 1954; Shotwell & Cattell, 1954; Stice & Cattell, 1954) needs little comment, except to say that it has recently been re-factored and independently rotated with confirmation of all but one of the factor positions (Cattell, 1955*a*). Besides general intelligence, it measures fifteen other factors, some of which are well known from clinical or general psychology; e.g. cyclothymia *v.* schizothymia (Factor *A*), emotional integration *v.* neuroticism (Factor *C*), surgency *v.* desurgency (Factor *F*), dominance *v.* submission (Factor *E*), etc., while others are new, e.g. sensitive emotionality *v.* toughness (Factor *I*) and self-sufficiency (Factor Q_2). Each factor score is based on ten or thirteen items in the *A* Form and the same number in the *B* Form, and the reliabilities of the factor scores (not to be confused with that of the items, by re-test) average 0·69. The whole test, which is untimed, takes from three-quarters of an hour to two hours.

Profiles, as means and sigmas for each factor, are already known on the 16 P.F. for the student population, the general adult population (Cattell, Stice & Saunders, 1950), and about twenty rather varied occupations. In addition to comparing the researcher's profile with these, we wished to compare it more sharply with those of teachers and administrators in the same areas, as indicated above. The areas taken were three: Physical Science (represented by Physics); Biological Science (represented by Biology as such); and Psychology (which, as Roe's results (1953) suggest, may be taken as representative to some extent of the social sciences).

Committees were set up to choose from the roster of their professional society members, especially the starred scientists (Visher, 1947; Cattell & Brimhall, 1921), between 200 and 250 individuals in each of the three areas. Since we were more interested in the researcher than in the others, they were instructed to choose about twice as many researchers as teachers and administrators (typically 100 R, 50 T, and 50 A). By definition the subjects were to be eminent or 'successfully engaged' in their profession, and specifically noted, in the one group, for research contribution, in the second for teaching (including text-book writing), but with low or negligible research publication (less than one article in 3 years), and, in the third, for administration (heads of departments, national committee organizers, board members, etc.), again with negligible comparable research contribution. For uniformity in an initial study we took American scientists, rather than British and American mixed, until such time as the test standardization should become wider in Britain. The numbers chosen and responding are shown in Table 1.

Table 1. *Samples of leading faculty members studied*

Later reference no. of group		No. designated	No. agreeing to co-operate	No. satisfactorily completing questionnaire	Mean and sigma of age of completers
	Biologists				
1	Teachers	62	29	29	44·7 (8·1)
2	Administrators	62	27	21	50·1 (9·1)
3	Researchers	93	55	46	48·7 (9·4)
	Subtotals	217	111	96	
	Physicists				
4	Teachers	50	28	25	43·0 (10·9)
5	Administrators	55	25	20	53·1 (8·3)
6	Researchers	121	53	46	51·4 (12·2)
	Subtotals	226	106	91	
	Psychologists				
7	Teachers	52	30	27	42·1 (9·0)
8	Administrators	48	33	28	48·0 (5·8)
9	Researchers	105	68	52	42·9 (8·1)
10	Selected researchers*	—	68	39*	43·2 (8·7)
	Subtotals	205	131	107	
	Grand total	648	348	294	

* This group of 39 was selected from the 52 by more exacting standards as to significance of research. It is interesting to see that in subsequent results it behaves 'like the research group (no. 9) only more so'. Not re-included in grand totals.

As usual we have no control on the selection by rejection, but it is at least satisfying that the percentage thus lost is surprisingly steady for different specialities and groups. We end, as we began, with about twice as many in the research group as in each of the other groups. The ages are also fairly uniform (the psychologists running somewhat younger) and high enough for the men to have had their work stably assessed by time. The selection of respondents by special committees rather than by simply taking, for example, starred men from *American Men of Science* (Visher, 1947), was required by various considerations, but chiefly by the need for separate ranking of general scientific faculty within the subgroups of teachers, administrators, and researchers.

The analysis of results is made in terms of means and significances of difference of means (first in raw scores, then in standard scores) on each factor. Analysis of variance would have no particular point here, and indeed our later purposes call simply for a basis for applying pattern similarity coefficients, or Q-technique, or multiple-discriminant functions, according to the present evidences as to which factors are significant for selection.

III. DIFFERENCES FROM THE GENERAL ADULT POPULATION

The raw score means and sigmas for each of the ten groups on all factors are shown in Table 2 (*a*). In Table 2 (*b*), which immediately follows, these values are converted into ten-point standard scores (mean = 5·5; sigma = 2 points) according to table 5 in tabular supplement A, for the Sixteen Personality Factor Handbook (Cattell *et al.* 1950). These norms are for a general adult population averaging 38 years, made by combining various occupational groups in proportions indicated by the census (Cattell *et al.* 1953). No age correction has been made, because knowledge of trends in the personality factors is still not accurate, and our groups are only a few years older (4–15 years) than the standardization group. Actually, the effect of the age trends as at present assessed is never more than 0·2 raw score point per year. In terms of standard scores, allowance for such trends would reduce slightly (about 1 point for 10 years) the levels in H and Q_3, and raise at about the same rate or less, the calculated levels in B, F, O, and Q_4, relative to those shown in Table 2 (*b*). Such a correction would apply to Biologists, A(dministrators) and R(esearchers); Physicists, A and R, and Psychologists, A and R′.

It will be seen at once that this group of leading scientists departs markedly from the pattern of the general population, and in a very consistent fashion, despite differences of field or mode of activity. They are in general more intelligent, more schizothyme, of greater ego-strength and dominance, less surgent, more adventurous, more sensitive, have less free-floating anxiety, and are slightly more radical and more self-sufficient. On super-ego strength, which is represented in a dual way, not yet entirely clear, by Factors G and Q_3, the evidence is conflicting. The Factor G, which represents emotional adoption of standard moral goals, is lower; but Factor Q_3, sometimes called 'will control' and which represents strict, internalized and intellectualized standards, and exacting demands on the self, is decidedly higher.

The reader may examine, from the sigmas in Table 2 (*a*) and the Standardization Table (Cattell *et al.* 1950), the significance of the 160 possible differences, but for economy of presentation we shall group the data into teachers (biologists, physicists, and psychologists together), administrators (similarly grouped), and researchers (similarly grouped), thus reducing to forty-eight comparisons. These are shown, as raw score differences with associated P values, in Table 3.

Thus, without attention to differences *among* teachers, administrators, and researchers, which are examined in the next section, we can see that the deviations from the general population listed from Table 2 (*b*) proved to be uniformly significant for all three groups.

Table 2 (*a*). *Raw scores. Means and sigmas for ten groups on all factors*

	A	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
Group 1. Biologists: T (29)																
Mean	11·79	20·41	34·03	23·03	18·90	23·10	27·76	18·83	17·24	19·24	24·14	12·21	25·55	25·76	27·48	13·93
Sigma	5·10	1·97	4·66	6·86	6·93	3·63	11·00	3·97	4·16	5·21	3·34	6·94	5·36	5·44	4·77	6·51
Group 2. Biologists: A (21)																
Mean	13·81	21·57	33·71	24·52	22·76	23·95	31·43	20·62	17·00	19·95	25·10	14·10	24·19	24·81	27·33	16·33
Sigma	5·24	2·82	5·09	6·12	5·64	4·97	7·66	3·71	5·08	4·77	4·89	8·12	4·07	6·85	4·84	8·11
Group 3. Biologists: R (46)																
Mean	10·02	21·43	31·96	22·41	18·35	23·91	27·54	20·87	16·50	21·24	24·30	13·26	25·87	28·22	27·48	14·52
Sigma	4·31	2·46	4·82	5·18	6·93	5·36	9·41	5·13	4·24	5·51	3·53	8·01	5·41	6·32	4·71	7·73
Group 4. Physicists: T (25)																
Mean	10·88	21·00	34·84	20·16	20·72	22·68	25·20	18·96	17·32	20·20	23·92	14·20	24·12	24·60	27·40	15·96
Sigma	4·15	2·50	4·65	5·69	8·67	6·00	10·31	3·75	3·65	5·64	3·95	6·04	5·09	5·72	5·68	7·54
Group 5. Physicists: A (20)																
Mean	8·65	20·45	35·35	20·00	18·50	26·00	28·00	17·50	15·75	15·95	24·70	13·25	21·65	25·80	29·25	13·25
Sigma	3·67	2·76	5·23	6·00	5·23	4·60	9·62	3·93	3·51	4·30	3·29	8·22	4·42	4·41	4·15	5·87
Group 6. Physicists: R (46)																
Mean	10·13	21·96	34·61	23·20	19·78	22·93	27·87	19·37	15·70	19·96	24·37	12·41	24·17	27·00	27·26	13·63
Sigma	3·99	2·47	4·14	5·78	6·96	6·26	10·04	3·96	4·62	7·25	4·10	7·62	4·27	4·10	5·21	7·17
Group 7. Psychologists: T (27)																
Mean	16·19	21·93	33·48	25·37	22·63	20·19	32·52	20·37	17·44	23·37	23·63	13·93	27·41	24·30	25·44	17·41
Sigma	4·79	2·40	6·59	6·26	6·29	5·59	8·55	4·91	5·54	5·42	3·85	8·25	4·97	4·88	4·41	9·13
Group 8. Psychologists: A (28)																
Mean	19·71	21·93	37·25	24·32	24·61	21·11	36·82	18·46	14·64	20·32	24·50	8·18	25·18	22·29	17·75	10·18
Sigma	5·14	2·00	4·12	5·17	6·64	5·60	7·60	4·73	4·08	4·41	2·91	7·12	4·57	5·46	4·90	6·03
Group 9: Psychologists: R (52)																
Mean	13·33	21·83	32·60	27·27	24·21	20·37	32·23	19·04	18·46	23·06	24·00	13·46	26·98	26·04	24·33	14·92
Sigma	5·29	2·87	5·98	5·78	6·83	4·77	9·78	4·43	4·84	6·34	4·45	7·20	5·23	6·74	6·39	6·90
Group 10. Psychologists: R′ (select) (39)																
Mean	13·26	21·67	32·38	26·72	23·21	20·33	30·92	18·54	18·05	22·44	23·95	13·26	26·95	26·51	24·26	14·92
Sigma	5·73	2·96	5·88	6·29	6·66	4·96	10·35	4·49	5·01	6·22	4·78	6·85	5·25	6·98	6·48	6·66

Table 2 (*b*). *Standard scores. Means for ten groups on all factors*

	A	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
Group 1. Biologists: T																
Mean	3·58	6·38	7·10	6·70	2·88	3·69	6·20	6·76	4·24	4·97	5·47	3·42	6·15	6·24	7·09	4·99
Group 2. Biologists: A																
Mean	4·44	6·73	7·03	7·28	4·08	4·02	6·93	7·43	4·14	5·23	5·91	3·94	5·79	6·02	7·05	5·66
Group 3. Biologists: R																
Mean	2·83	6·69	6·66	6·47	2·70	4·00	6·16	7·53	3·94	5·70	5·55	3·71	6·24	6·82	7·09	5·15
Group 4. Physicists: T																
Mean	3·19	6·56	7·27	5·60	3·44	3·53	5·70	6·81	4·27	5·32	5·37	3·97	5·77	5·97	7·07	5·54
Group 5. Physicists: A																
Mean	2·24	6·39	7·38	5·54	2·75	4·81	6·25	6·25	3·64	3·77	5·73	3·71	5·11	6·25	7·54	4·80
Group 6. Physicists: R																
Mean	2·87	6·85	7·23	6·77	3·15	3·63	6·23	6·96	3·62	5·23	5·58	3·48	5·79	6·54	7·03	4·91
Group 7. Psychologists: T																
Mean	5·45	6·84	6·99	7·60	4·04	2·57	7·15	7·34	4·32	6·47	5·24	3·89	6·65	5·90	6·57	5·94
Group 8. Psychologists: A																
Mean	6·95	6·84	7·79	7·20	4·66	2·93	8·00	6·62	3·20	5·36	5·64	2·32	6·05	5·43	7·16	3·96
Group 9. Psychologists: R																
Mean	4·24	6·81	6·80	8·33	4·53	2·64	7·09	6·84	4·72	6·36	5·41	3·76	6·53	6·31	6·29	5·26
Group 10. Psychologists: R′																
Mean	4·21	6·76	6·75	8·12	4·22	2·63	6·83	6·65	4·56	6·13	5·39	3·71	6·53	6·42	6·27	5·26

In intelligence, ego strength, dominance, desurgency, lack of group super-ego, adventurousness, sensitive emotionality, freedom from paranoid trends, freedom from free-floating anxiety, and higher 'internalized compulsive super-ego' (formerly labelled 'will control') the differences for all three groups are significant away beyond the 1% level. In schizothymia, radicalism, and self-sufficiency the uniformity is broken by the Administrator group's divergence falling to only the 1–10% significance level.

It is interesting to compare these deviations of the 'genus' *successful academic man*, as measured from the 'order' *general population*, also with their deviations from the 'family' *general university student*, as shown by undergraduate norms. Due to the large age difference that would here be involved, and the impossibility of accurately bridging it with present knowledge of personality factor life changes, no table has been set out. However, from table 3 in Supplement A of the 16 P.F. Handbook (university undergraduate students, men and women), and using the allowances for age trends over 25 years, given in table 6 therein, we can at least state indications. With the student profile thus projected forward 25 years the outstanding faculty now show no difference on ego strength (C), dominance (E), super-ego: group standards (G), sensitive emotionality (I), free anxiety (O), and internalized super ego (Q_3). They switch over on adventurousness (H) from high to somewhat low, on paranoid trend (L) from low to somewhat high, and on psychosomatic anxiety (Q_4) from slightly low to slightly high. They still deviate, almost as much as before, in the direction of being more intelligent (B) and more radical (Q_1). Finally, in schizothymia ($A-$) and self-sufficiency (Q_2) they deviate as they did from the general population, only more so!

The only uncorrected influence of unknown magnitude in the above is from the larger proportion of males (see Table 1) in the outstanding faculty than in the standardization groups for students and adult occupational composites. However, surprisingly few significant differences have been found between male and female populations, either in the form or the level of their general personality factors. The only observed differences significant at the 1% level (Cattell *et al.* 1950, 1953) are for women to be more cyclothyme ($A+$), less dominant ($E-$), and more sensitively emotional ($I+$). On this basis, if more balanced for sex, our outstanding academics might deviate less in schizothymia and dominance, and show some slight deviation from the student group in the same direction as from the general population in Factor I, toward sensitive emotionality. If entirely male norms were substituted for mixed norms in our control general population and student groups the deviation of academics on these three sex-modifiable factors would still be in the direction of greater schizothymia, dominance, and sensitive emotionality, regarding the general population, but dominance would cease to have significant deviation from male students.

These differences of pattern against different backgrounds may become confusing to the reader without time to study the various tables and corrections, so in conclusion we will concentrate on the contrast with the student type. Here it is interesting to observe that the deviations of these successful professional men are in agreement with those found by Wright (1954) to distinguish men of equal intelligence but higher scholastic achievement from those of lower achievement. His differences significant at $P=0{\cdot}01$ were: greater schizothymia ($A-$) and withdrawn schizothymia ($H-$); greater ego strength (C), absence of free-floating anxiety ($O-$) and of somatic anxiety (Q_4-). The three last do not distinguish our men from students, but do distinguish them from the general population. To

Table 3. Differences of means of teachers (T), administrators (A), and researchers (R), from general population

(Positive scores indicate higher than the general population in the usual direction of measurement of the factor, i.e. that of the name on the left of the bipolar title.)

	T		A		R	
	Diff.	*P*	Diff.	*A*	Diff.	*P*
A. Cyclothymia *v*. schizothymia	−3·32	<0·001	−1·59	0·03	−5·05	<0·001
B. Intelligence *v*. defective general ability	3·60	<0·001	3·89	<0·001	4·24	<0·001
C. Ego Strength *v*. general emotionality	7·60	<0·001	9·12	<0·001	6·53	<0·001
E. Dominance *v*. submissiveness	3·03	<0·001	3·23	<0·001	4·52	<0·001
F. Surgency *v*. desurgency	−6·60	<0·001	−5·20	<0·001	−6·38	<0·001
G. Super-ego, identification with group standards *v*. lack of super-ego	−5·80	<0·001	−4·41	<0·001	−5·48	<0·001
H. Adventurous cyclothymia *v*. withdrawn schizothymia	4·36	<0·001	8·42	<0·001	5·14	<0·001
I. Sensitive emotionality *v*. toughness	3·88	<0·001	3·34	<0·001	4·23	<0·001
L. Paranoid trend *v*. lack of paranoid trend	−3·07	<0·001	−4·72	<0·001	−3·45	<0·001
M. Bohemian unconcern	0·21	0·37	−1·76	<0·01	0·79	0·08
N. (Psychological title uncertain)	−0·30	0·25	0·54	0·13	0·02	0·48
O. Free-floating anxiety *v*. lack of anxiety	−6·40	<0·001	−8·35	<0·001	−6·74	<0·001
Q_1. Radicalism *v*. conservatism	2·63	<0·001	0·76	0·10	2·63	<0·001
Q_2. Self-sufficiency *v*. lack of self-sufficiency	2·31	<0·01	1·47	0·02	4·44	<0·001
Q_3. Super-ego, will control *v*. lack of will control	5·58	<0·001	6·86	<0·001	5·07	<0·001
Q_4. Psychosomatic anxiety	−0·08	0·46	−2·84	<0·01	−1·42	0·01

Table 4. *Means and sigmas for teachers, administrators, and researchers on all factors*

	A	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
Teachers (81)																
Mean	12·98	21·10	34·10	22·93	20·70	22·00	28·56	19·38	17·33	20·91	23·90	13·40	25·73	24·91	26·78	15·72
Sigma	5·20	2·35	5·34	6·58	7·39	5·22	10·35	4·25	4·47	5·64	3·67	7·12	5·26	5·32	4·99	7·82
Administrators (69)																
Mean	14·71	21·39	35·62	23·13	22·28	23·39	32·62	18·84	15·68	18·94	24·74	11·45	23·86	24·07	28·06	12·96
Sigma	6·61	2·54	4·93	5·98	6·40	5·47	8·95	4·33	4·31	4·83	3·67	8·12	4·56	5·78	4·67	7·09
Researchers (144)																
Mean	11·25	21·74	33·03	24·42	20·92	22·32	29·34	19·73	16·95	21·49	24·22	13·06	25·73	27·04	26·27	14·38
Sigma	4·83	2·61	5·17	5·97	7·31	5·64	9·92	4·57	4·71	6·49	4·04	7·56	5·10	5·91	5·68	7·23

Wright's picture our study adds that the scientific 'achievers' are higher in intelligence (B), radicalism (Q_1), self-sufficiency (Q_2), and perhaps sensitive emotionality (I) and paranoid trend (L). In personality they are thus not simply 'academic over-achievers', and indeed, considered in the light also of their profile deviations from the general population, which is the more basic comparison, they are distinguished, in addition to the above 10 factors, by high dominance (E), high internalized super-ego or will control (Q_3) and marked desurgency ($F-$).

It is to be hoped that these findings on well known factors will lead to fuller understanding of the psychological conditions of and selection pathways toward scientific eminence, as well as to additional insight and predictive power in regard to the psychological nature of the personality factors themselves.

IV. THE DIFFERENCES OF RESEARCHERS FROM ADMINISTRATORS AND TEACHERS

Having defined the genus of 'outstanding academic scientists', we can now discover within it the characters of the particular species—the researcher—in which we are primarily interested.

Table 4 sets out the personality factor means and sigmas for the three groups (from which the values both of Table 3 and Table 5 are derived).

The significance of differences of means among the Table 4 groups is shown in Table 5. The P values are for a two-tailed distribution, i.e. assuming no hypothesis has been made as to the direction of difference, though actually hypotheses in some cases *were* made (Cattell, 1955*b*) and the P values might be halved in those instances.

Table 5. *Significance of differences among teachers (T), administrators (A), and researchers (R)*

(Positive differences indicate higher mean scores in the category mentioned second in each comparison; measurement in the same direction as in Table 3.)

	T *v.* A		T *v.* R		A *v.* R	
	Diff.	*P*	Diff.	*P*	Diff.	*P*
A. Cyclothymia-schizothymia	1·73	0·08	−1·73	0·01	−3·46	0·001
B. Intelligence	0·29	0·47	0·64	0·07	0·35	0·36
C. Ego strength	1·52	0·07	−1·07	0·15	−2·59	0·001
E. Dominance	0·20	—*	1·49	0·09	1·29	0·15
F. Surgency	1·58	0·16	0·22	—	−1·36	0·17
G. Super-ego	1·39	0·11	0·32	—	−1·07	0·18
H. Adventurous cyclothymia	4·06	0·01	0·78	—	−3·28	0·02
I. Sensitive emotionality	−0·54	0·44	0·35	—	0·89	0·17
L. Paranoia	−1·65	0·02	−0·38	—	1·27	0·07
M. Bohemian unconcern	−1·97	0·02	0·58	0·46	2·55	0·001
N. (Psychological title uncertain)	0·84	0·23	0·32	—	−0·52	0·36
O. Free anxiety	−1·95	0·12	−0·34	—	1·51	0·19
Q_1. Radicalism	−1·87	0·02	0·00	—	1·87	0·01
Q_2. Self-sufficiency	−0·84	0·37	2·13	0·01	2·97	0·001
Q_3. Super-ego, will control	1·28	0·11	−0·51	0·48	−1·79	0·02
Q_4. Psychosomatic anxiety	−2·76	0·02	−1·34	0·21	1·42	0·17

* Any P value beyond 0·50 is represented by a —.

Suggestive differences are visible in the tendency of researchers to have higher intelligence (B) than teachers ($P=0{\cdot}07$), to be more dominant (E) than either teachers ($P=0{\cdot}09$) or administrators, and to be more paranoid (L) than administrators ($P=0{\cdot}07$). Differences immediately acceptable at the $P=0{\cdot}05$ to $0{\cdot}01$ level are: the tendency to more withdrawn schizothymia ($H-$) and to lower compulsive super ego (Q_3) than administrators. Any difference is a contribution to definition of both parties, and some of the above are as much comments on the more outgoing qualities of administrators as on the more withdrawing qualities of researchers; but finally we shall stress the definition of the researcher in terms of the ways he differs simultaneously from administrators and teachers.

Thus, at or beyond the 1% level of certainty he is more schizothyme ($A-$) and more self-sufficient (Q_2) than teachers or administrators. At the same level he shows more emotional instability ($C-$) and Bohemian unconcern (M) than administrators, but in this case teachers are similar to the administrator block at a less clear P level. Also beyond the 1% level is the researcher's tendency to be more radical (Q_1) than the administrator.

Table 6. *Significance of differences among teachers, administrators and researchers for separate scientific areas*

Read as for table 5; italicized figures refer to 'Selected researchers' (R') (see Table 1).

	T *v.* A		T *v.* R		A *v.* R		R *v.* R′	
	Diff.	*P*	Diff.	*P*	Diff.	*P*	Diff.	*P*
Biologists								
H. Adventurous cyclothymia	3·67	0·08	−0·22	0·46	−3·89	0·04		
I. Sensitive emotionality	1·79	0·05	2·04	0·03	0·25	0·41		
L. Paranoia	−0·24	0·43	−0·74	0·23	−0·50	0·35		
M. Bohemian unconcern	0·71	0·31	2·00	0·06	1·29	0·16		
Q_3. Super-ego, will control	−0·15	0·46	0·00	0·50	0·15	0·45		
Q_4. Psychosomatic anxiety	2·45	0·23	0·59	0·36	−1·86	0·19		
Physicists								
H. Adventurous cyclothymia	2·80	0·17	2·67	0·15	−0·13	0·48		
I. Sensitive emotionality	−1·46	0·10	0·41	0·33	1·87	0·04		
L. Paranoia	−1·57	0·07	1·62	0·05	−0·05	0·48		
M. Bohemian unconcern	−4·25	<0·01	−0·24	0·44	4·01	<0·01		
Q_3. Super-ego, will control	1·85	0·10	−0·14	0·46	−1·99	0·05		
Q_4. Psychosomatic anxiety	−2·71	0·09	−2·33	0·10	0·38	0·41		
Psychologists								
H. Adventurous cyclothymia	4·30	0·02	−0·29	0·44	−4·59	0·01		
			−1·00	*0·25*	*−5·90*	*<0·01*	*−1·31*	*0·27*
I. Sensitive emotionality	−1·91	0·07	−1·33	0·44	0·58	0·01		
			−1·83	*0·25*	*0·08*	*<0·01*	*−0·50*	*0·27*
L. Paranoia	−2·80	0·02	1·02	0·21	3·82	<0·001		
			0·61	*0·32*	*3·41*	*<0·01*	*−0·41*	*0·35*
M. Bohemian unconcern	−3·05	0·01	−0·31	0·41	2·74	0·01		
			−0·93	*0·26*	*2·12*	*0·05*	*−0·62*	*0·32*
Q_3. Super-ego, will control	2·31	0·03	−1·11	0·18	−3·42	<0·01		
			−1·18	*0·19*	*−3·49*	*0·01*	*−0·07*	*0·48*
Q_4. Psychosomatic anxiety	−7·23	0·01	−2·49	0·11	4·74	<0·01		
			−2·49	*0·11*	*4·74*	*<0·01*	*0·00*	*0·50*

Since some differences may be partly peculiar to certain scientific areas, and thus may be masked by arranging all areas together, about one-half of the total possible area breakdowns—those of sufficient interest to justify space—are presented in Table 6. (All have been examined and the full table is presented elsewhere (Cattell, 1955*b*)). In general we may note a tendency for differences to be more significant among psychologists, reflecting either greater specialization or greater skill of that selection committee (motivated by both scientific and psychological interests) in sorting out the criterion types. There is also a tendency for the more selected psychological research group, R′, to show more significant differences than the larger group, R.

The agreement of the differences for the subgroups in contributing to those 0·01 level differences for the total group cited from Table 5 is so uniform that only one of those five factors needs to be relisted here in order to inspect an inconsistency in the parts. This is M, Bohemian unconcern, where it will be seen that there is an inversion in psychology, teachers being slightly more 'Bohemian' ($M+$) than researchers. Factor E, not reproduced here, also shows an inversion of the usual greater dominance of researchers in the special case of biology, and these biological researchers are also significantly ($P = 0{\cdot}01$) more desurgent ($F-$) than their colleagues, which suggests perhaps aesthetic rather than mastery interests in much biological research.

Turning to the differences that are of rather low significance in the composite picture and examining them in individual areas we note first that the higher 'paranoid trend' score is really peculiar to psychological researchers, in whom it reaches $P < 0{\cdot}014$ and $< 0{\cdot}001$. Possibly, in the more chaotic state of a new science, some tenacity of conceptualization is a necessary escape from despair! And perhaps those more interested in nature's laws than man's theories choose a more developed, structured science than psychology. The withdrawn schizothymia factor, H, is perceived in Table 6 to be no longer selective of researchers from teachers in the case of physics, but to be highly selective of psychological researchers, especially in the purer research group (R′). The effect is partly due to psychologist *teachers* being more adventurous cyclothymic ($H+$) than most. Presumably teaching in an area in which research has yet provided so little of an exact nature to talk about attracts a more sociable, adventurous type, perhaps more interested in the interactive process of teaching than in the subject-matter.

Factor I is included in Table 6 because it shows in each group a uniform tendency toward significance which is nevertheless missed in the composite picture. Researchers, relative in one case to administrators and another to teachers, tend to be more sensitively emotional ($I+$). Only the psychologists upset this. The trend is most clearly shown in the biologists, again suggesting aesthetic-emotional involvements in the motivation of biological research.

Factor M (Bohemian unconcern) is set out because the breakdown shows that the researchers' deviation is largely due to administrators being distinctly low in it. The latter show marked 'group concernedness' and indeed in physics and psychology the teachers are higher on unconcern than the researchers. Factor Q_3, internalized compulsive super-ego or will control is similar, in that the difference is principally due to the administrators showing the high regard for exactness, high standards of behaviour and rather obsessional trends which form the essence of the factor, though teachers are also higher than researchers. The same form of relation holds for Q_4, the absence of somatic anxiety being characteristic of administrators rather than its presence characteristic of researchers.

V. SUMMARY AND DISCUSSION

1. The personality profile on the 16 P.F. Test has been compared for larger groups of scientists (96 biologists, 91 physicists, and 107 psychologists) than in any previous investigation of the personality of the researcher, and several differences significant at the $< 0{\cdot}01$ or $< 0{\cdot}001$ levels have been discovered: (*a*) between scientists and: (i) the general population, (ii) the student population; and (*b*) between researchers, teachers and administrators, all in the field of science.

2. The outstanding professional, academic scientists as a whole differ from:

(*a*) *The general population*, at beyond the 1 % level, in the personality factors called general intelligence (B), ego strength or stability (C), dominance (E), desurgency ($F-$), lack of group super-ego standards ($G-$), adventurousness (H), sensitive emotionality (I), lack of paranoid trends ($L-$), lack of free-floating anxiety ($O-$), and compulsive super-ego (or will control) (Q_3). Except for administrators they are at the same significantly higher level in radicalism (Q_1) and self-sufficiency (Q_2). Administrators alone differ at the 1 % level in showing lower somatic anxiety (Q_4-).

(*b*) The university *undergraduate population*, approximately age corrected, by being decidedly more schizothyme ($A-$) and self-sufficient (Q_2), more intelligent (B), desurgent ($F-$), withdrawn schizothyme ($H-$), radical (Q_1) and, probably, more paranoid (L) and of higher somatic anxiety (Q_4).

3. Researchers, relative to teachers or administrators, are, at the 1 % level of certainty, more schizothyme ($A-$), self-sufficient (Q_2), emotionally unstable ($C-$), bohemianly unconcerned (M), and radical (Q_1). They are also significantly, but less uniformly, more dominant (E), paranoid (L), withdrawn schizothyme ($H-$) and lower on compulsive super-ego (will control) (Q_3). In the last and in radicalism they differ more from administrators than from teachers. There are some interesting differences indicated among the sciences, researchers in biology being, among biologists, less dominant and more desurgent; but these inter-area differences are more fully discussed elsewhere (Cattell & Drevdahl, 1955).

These results may be compared with Drevdahl's study (1954) of 'creative' *v.* 'non-creative' students in science and in the arts, which used the same personality factors and the same measuring instrument. The creative were significantly more schizothyme ($A-$), self-sufficient (Q_2), withdrawn schizothyme ($H-$), sophisticated ($N+$), desurgent ($F-$), and radical (Q_1). There is thus important agreement, while there are also some special characteristics of creative persons at the student level which do not particularly show up in those called to research (and vice versa).

Certain hypotheses had been reached from reading biographies of one hundred eminent British, American and other researchers prior to these experimental determinations (Cattell, 1955*b*). These personal records suggested that researchers would be more schizothyme, intelligent, dominant, paranoid, self-sufficient, and of greater super-ego development than the generality of professions of the same educational level. These characteristics of past researchers seem well matched by our findings on living researchers, except for super-ego strength, where the difference tends to be, if anything, reversed.

Brevity of space compels deferment of discussion of the psychological and social operation of these personality factors to another place (Cattell, 1955*b*). But it is easy to see that the schizothymic preoccupation with things and ideas, rather than people; the self-sufficiency which favours creativity and independence of mind; the dominance which gives satisfaction in mastery of nature for its own sake; and the emotional instability which permits radical re-structuring and creativity, would all be vital to the best kind of basic research performance—though perhaps unpleasant in an administrator and inapt in a business man. The remaining factors of Bohemian unconcern and radicalism may perhaps best be regarded as condition and consequence of continued adjustment to the researcher's comparative earnings!

It is suggested that until further research, especially with the improved personality factor measures now becoming available (Cattell, 1954, 1955*a*), is prosecuted on this problem any vocational selection in progress be based on the standard score profile derived from row 5 of Table 4, and which is set out in standard form in Table 7. Naturally it should be used with due regard to its tentative character and in conjunction with whatever other evidence is available for selecting researchers from examination-passers at the university level.

Table 7. *Profile of the existing typical researcher* (*in general population standard scores*)

Factor		
A	Cyclothymia *v.* schizothymia	3·34
B	Intelligence	8·65
C	Stability	6·89
E	Dominance	7·24
F	Surgency	3·50
G	Super-ego, group standards	3·40
H	Adventurous cyclothymia	6·44
I	Sensitive emotionality	7·1
L	Paranoia	6·92
M	Bohemianism	5·78
N	—	5·50
O	Free-floating anxiety	3·66
Q_1	Radicalism	6·20
Q_2	Self-sufficiency	6·52
Q_3	Super-ego, will control	6·78
Q_4	Psychosomatic anxiety	5·11

As in other I.P.A.T. tests the 16 P.F. Standard Score is arranged in *Stens* not Stanines, i.e. each point is $\frac{1}{2}$ a sigma and the population mean is 5·5, so that 1 and 10 are equidistant from the mean and 5·0 to 5·99 is essentially an average score.

This personality profile reminds us, after the above comparisons of the genus 'scientist' with the average, and the species 'researcher' with other specialists, of the essential divergence of the species 'researcher' from the average man. For example, though on the unstable side ($C-$) relative to other professionals in science he is yet decidedly above general, average stability. And though no more desurgent than other specialists he is yet emphatically below the general adult level of surgency.

Now the use of the pattern similarity coefficient, r_p (Cattell, 1949; Cronbach & Gleser, 1953) to determine the congruence of various propositi to the above profiles, when selecting promising men for research openings, involves certain assumptions which should be watched. First it gives equal weight to the various factors and secondly it assumes that the values for the existing researcher are the best values. While the first is the best guess with our present knowledge, and the second must be accepted unless we can claim to be wiser than society's laws of supply and demand,[1] yet we can hope to find out a little more. Just as the first step in vocational selection research is to define the type profile, so the second is to determine the linear or curvilinear relation of each factor to a criterion of success in a population *within* the occupation.

A beginning has fortunately been made by Van Zelst & Kerr (1954) with this second step, with results consonant with our own; but it needs tackling on an adequate scale and with radical attention to the criterion problem, perhaps along lines indicated elsewhere (Cattell, 1955*b*).

---

[1] As a psychologist, one may clearly realize that Factor B, intelligence, is likely to produce indefinite improvement of performance with higher values, but it is necessary also to be a sociologist and ask whether it is correct to advise a person to enter an occupation when his I.Q. is higher than that of persons at present normally attracted to, and perhaps more vitally needed in some other occupation.

REFERENCES

CATTELL, J. McK. & BRIMHALL, D. R. (1921). *American Men of Science*, 3rd ed. Garrison, New York: Science Press.

CATTELL, R. B. (1949). r_p and other coefficients of pattern similarity. *Psychometrika*, XIV, 279–98.

CATTELL, R. B. (1950*a*). *Personality: a Systematic Theoretical and Factual Study*. New York: McGraw Hill.

CATTELL, R. B. (1950*b*). The principal culture patterns discoverable in the syntal dimensions of existing nations. *J. Soc. Psychol.* XXXII, 215–53.

CATTELL, R. B. (1950*c*). The main personality factors in questionnaire self estimate material. *J. Soc. Psychol.* XXXI, 3–38.

CATTELL, R. B. (1954). *The O-A (Objective-Analytic) Personality Test Battery.* I.P.A.T. 1602 Coronado Drive, Champaign, Illinois.

CATTELL, R. B. (1955*a*). A shortened 'basic English' version (Form C) of the 16 P.F. questionnaire. I.P.A.T. 1602, Coronado Drive, Champaign, Illinois.

CATTELL, R. B. (1955*b*). *Research: the Task and the Man* (in the Press).

CATTELL, R. B., DAY, M. & MEELAND, T. (1953). La standardisation du questionnaire de personalité en 16 facteurs de l'I.P.A.T., *Rev. Psychol. appl.* III, no. 2.

CATTELL, R. B. & DREVDAHL, J. E. (1955). Personality differences among eminent biologists, physicists and psychologists, as determined by the 16 P.F. test (in preparation).

CATTELL, R. B., STICE, G. F. & SAUNDERS, D. R. (1950). *The Sixteen Personality Factor Test and Handbook.* I.P.A.T. 1602 Coronado Drive, Champaign, Illinois.

CRONBACH, L. J. & GLESER, G. C. (1953). Assessing similarity between profiles. *Psychol. Bull.* 50, 456–73.

DREVDAHL, J. E. (1954). An exploratory study of creativity in terms of its relationships to various personality and intellectual factors. Ph.D. Thesis, Library of University of Nebraska, Lincoln, Nebraska.

GUILFORD, J. P., WILSON, R. C. & CHRISTENSEN, P. R. (1952). A factor-analytic study of creative thinking. II. Administration of tests and analysis of results. *Rep. Psychol. Lab.* no. 8. Los Angeles: University of Southern California.

KRETSCHMER, E. (1931). *The Psychology of Men of Genius.* London: Kegan Paul.

ROE, A. (1953). A psychological study of eminent psychologists and anthropologists, and a comparison with biological and physical scientists. *Psychol. Monogr.* APA, Washington, D.C., 67, no. 2. (Whole no. 352).

SHOTWELL, A. & CATTELL, R. B. (1954). Personality profiles of more successful and less successful psychiatric technicians. *J. Ment. Def.* LVIII, 3.

SPEARMAN, C. (1933). *Creative Ability.* London: Nisbet.

STICE, G. F. & CATTELL, R. B. (1954). Four formulae for selecting leaders on the basis of personality. *Hum. Relat.* (in the Press).

THORNDIKE, R. L. & NORRIS, R. C. (1952). Empirical evidence on Air Force career fields. *Res. Bull. USAF Hum. Resour. Res. Cent.*, 52–13.

VAN ZELST, R. H. & KERR, W. A. (1954). Personality self-assessment of scientific and technical personnel. *J. Appl. Psychol.* XXXVIII, 145–7.

VISHER, S. S. (1947). Starring in *American Men of Science*, CVI, 359–61.

WOLFE, D. (1952). *Human Resources: The Needs and the Supply.* Pittsburgh: University of Pittsburgh Press.

WRIGHT, S. (1954). Some psychological and physiological correlates of certain academic underachievers. Ph.D. Thesis, University of Chicago.

(*Manuscript received* 15 *November* 1954)

SUBJECTIVE JUDGEMENTS—SOME EXPERIMENTS EMPLOYING EXPERIENCED AND INEXPERIENCED OBSERVERS[1]

By R. G. HOPKINSON

Building Research Station, Garston, Watford, Herts

I. *Introduction* (p. 262). II. *The performance of experienced and inexperienced subjects* (pp. 262–264). III. *Experimental studies of the effect of experience* (pp. 264–267). IV. Stability of the observer (pp. 267–269). V. *The selected team and the general population* (pp. 269–270). VI. *Discussion* (pp. 270–272). VII. *Summary* (p. 272). *References* (p. 272).

I. Introduction

Design of equipment for use by human beings needs to be closely related to the detailed demands of the users. The human being himself can be the only satisfactory judge of the efficiency of the design. This fact has never been completely ignored, but there has been a tendency to direct most effort to the perfection of the physical measurement of design efficiency and to disregard the fact that to be of any real value measurement must be related to the judgement of the human being who is to use the equipment.

Recently there has been a healthy sign that this mechanistic attitude is giving way to other methods. Designs of equipment are tested systematically by people and not by physical measures. Undoubtedly one of the main causes of this change in outlook has been the introduction of simple statistical methods which enable subjective experiments to be designed soundly, and their results interpreted by a strict comparative statistical procedure, rather than by inspection and induction alone.

With these refinements in the technique of subjective judgement experiments has come realization that, if human beings are to be used as meters, they must be properly calibrated. The techniques used for calibration of human meters have revealed that some human meters drift in their calibrations with time. The period over which these drifts take place is often very long, perhaps months, and so quantitative evidence is hard to come by.

During the course of some lengthy studies on glare-discomfort, quantitative evidence of the behaviour of human meters has been collected which has been thought worth recording, particularly as this evidence seems to confirm well-established opinion. Quantitative evidence is still lacking for a satisfactory statistical treatment, but it has been considered worth while to put forward what is available at this stage because many workers in the field of subjective judgements appear to be unaware of these tendencies on the part of their subjects. These tendencies may influence their results and so lead to significant errors.

II. The performance of experienced and inexperienced subjects

The performance of subjects making mental assessments of their environment seems to be related to the experience of the subject of the particular phenomenon under study. It is unfortunately not too easy to define exactly what is meant by experienced and inexperienced subjects. For the purposes of the discussion an experienced subject is considered to be one who is not only familiar with the technique of making mental assessments under experimental conditions, but who is also familiar with the nature of the problem being investigated. For example, an experienced subject in an experiment for the judgement of picture quality on a television screen, would be one who is not only familiar with the technique of judging the quality of photographs, that is to say, tone range, contrast, definition and so on, but who is also familiar with the television problem itself, who has made judgements of a number of different systems and is aware of the nature and importance of the variables being studied. An inexperienced observer is one who has not these advantages.

---

[1] Paper read at the meeting of the British Association for the Advancement of Science, Belfast, September 1952.

The main experiments referred to in this paper are concerned with the judgement of glare discomfort in buildings. The background to the work, however, extends back many years and includes detailed subjective judgement experiments on the assessment of the quality and luminosity of photographs for true representation, the judgement of ease of reading under a wide range of illumination and contrast conditions, the assessment of glare discomfort in lighted streets, various wartime visibility problems such as the assessment of the visibility of radar signals under a wide variation of display types, and so on.

Three conclusions emerged very early in the attempts at systematic subjective judgements.

(1) Subjects with experience of the phenomena under investigation gave observations with a greater repetition accuracy than those of inexperienced subjects.

(2) The mean of observations from subjects selected at random from the population was sometimes, but not always, significantly different from the mean of observations from skilled observers.

(3) The diversity of opinion between inexperienced observers was often greater than that between experienced observers.

A considerable amount of data led to the belief that, if reasonably speedy progress were to be made in an experiment involving subjective assessments, it would be preferable every time to build up a small panel of five or six skilled observers to make the judgements, rather than to hope for anything useful to be achieved in a reasonable time with observers selected at random from the population.

During the war when speed was essential, this technique of relying on a small team of skilled observers was fully justified. Care had to be taken, however, to use the team only for those types of experiments where it could be reasonably certain that the mean values of the observations given by the small team would be sensibly the same as those given by the population who were going to use the particular piece of equipment under test.

There was evidence right from the start that in certain types of judgement the skilled team gave observations different from the general population, although this was by no means always the case. The distinction between the observations of inexperienced and experienced subjects depended on the nature of the phenomena under investigation.

For example, in the estimation of discomfort, the experienced observer would recognize consistently the existence of a sensation of discomfort under conditions which the inexperienced observer considered to be satisfactory. On the other hand, in the estimation of visibility the experienced observer would be able to recognize detailed characteristics of a signal consistently under conditions which would barely permit the inexperienced observer to observe its presence. With both these types of experiment the distinction between the experienced and the inexperienced observers shows itself up by differences in their mean assessments as well as in the variability of their observations.

In other experiments however, for example, the grading of contrast scales in steps, no difference between the mean of the observations given by the experienced and the inexperienced observers shows itself. The distinction is apparent only in the much greater variance of the inexperienced observer's judgements.

III. EXPERIMENTAL STUDIES OF THE EFFECT OF EXPERIENCE

An investigation of the effect of various physical factors on the sensation of glare discomfort in the lighting of buildings afforded an opportunity of studying the effect of experience on the sensitivity to discomfort. This study made use of the Multiple Criterion technique which includes a series of 'calibration' observations over a long period of time. The technique also demands the freedom on the part of the subject to change without restraint the physical conditions governing the situation in which he finds himself. This enables him to make settings corresponding to a predetermined series of criteria. Details of the technique have been described elsewhere (Hopkinson, 1950), but a brief recapitulation may be useful.

The essence of the Multiple Criterion technique is to present the subject with a stimulus situation in which he has control of one of the significant physical variables whose change influences the magnitude of the sensation which is being studied. No effort is made, however, to ask the subject to put any sort of numerical scale on the magnitude of the sensation. He is at first given the definitions of the series of related criteria of the sensation covering, if possible, the useful gamut of the sensation in not more than four distinct steps. He is then asked to adjust the magnitude of the physical variable to correspond, in turn, with each of this limited number of defined related criteria corresponding to the subjective response. In this way it is found that each criterion acts as a check upon judgements made in terms of the others. It gives the subject greater confidence in making his judgement. Differences between subjects in the setting of the physical variables for the same criterion of sensation should also be less attributable to differences in the interpretation of the description of the sensation.

A further essential feature of the technique is to have a 'control experiment' extending over the whole period of the investigation. This is effectively a calibration from day to day of the human being as a meter. The subject is presented with a series of standard situations in each of which everything is constant except one physical variable. He is then asked to adjust this particular physical variable to correspond in turn to a predefined series of criteria of sensation. This judgement is repeated with each of the situations. The same situations are presented to him day after day, and settings of the same control experiment permit an analysis to be made of the scatter of the subject's observations under standard conditions, so that confidence limits can be derived for other observations under similar conditions. It also permits any effect of the influence of experience to be detected by a progressive drift in the observations. It is therefore chiefly with the observations obtained under the 'controlled experiment conditions' that this discussion deals.

Fig. 1 shows a typical set of observations of such a 'controlled' experiment. The assessment here is of the surround brightness with which a given glare source gives 'just uncomfortable' glare discomfort (Petherbridge & Hopkinson, 1950). The surround brightness is the physical variable under the observer's control, and he adjusts this to balance out the glare discomfort from the glare sources, which are fixed in size, position and brightness while the judgement is made. This observer is 'experienced' in the sense that he is a skilled photometrician and is familiar with the glare problem and with the variables that control it.

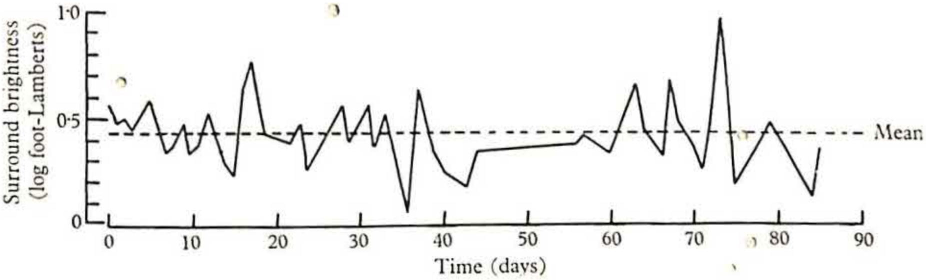


Fig. 1. Set of 'Control Experiment' observations showing an *experienced* observer's day-to-day assessments of the surround brightness with which given glare sources give 'just uncomfortable' glare. (Source brightness, 1000 ft.L.; observer P.P.)

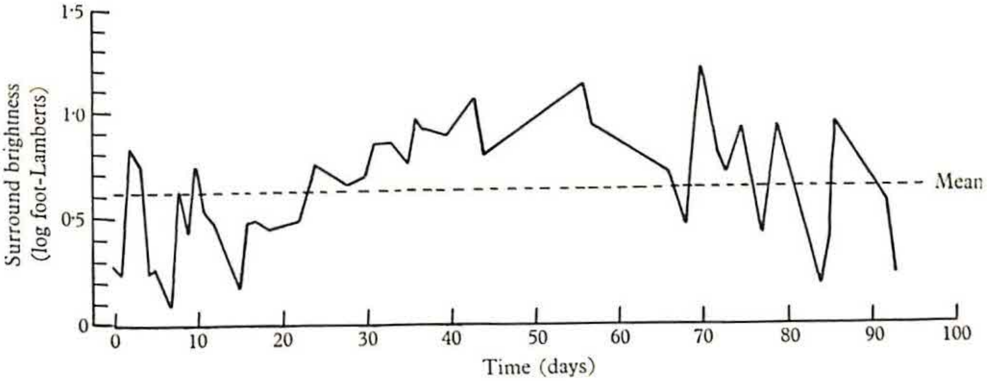


Fig. 2. Set of 'Control Experiment' observations given by an *inexperienced* observer. (Source brightness, 1000 ft.L.; 'just uncomfortable' glare; observer J.L.)

Throughout a period of nearly 90 days (except for a break of a fortnight in the middle of the series) observations made almost daily show no progressive trend. There are occasional observations which differ widely from the mean but, apart from these, the variance is quite small when it is remembered that the only guide which the observer has in making these observations is the degree of glare discomfort which he experiences. He has no means of recognizing any feature of the situation, and so cannot make his settings by some form of memory match.

These and similar data suggest fairly strongly that experienced observers who have formed a criterion of a sensation under controlled conditions can hold that criterion over a very long period. Considerable random variations will occur but the mean of a sufficiently large number of observations will remain stable.

A different situation appears to hold for inexperienced observers. Fig. 2 (observer J.L.) shows a set of observations taken by an observer who at the time had no experience of brightness judgements and of the glare problem. It can be seen that there has been a steady increase in the sensitivity to glare as the observer gains experience of the control experiment. (Other experiments in which the subject acted as an observer ran concurrently with the 'control experiment', thus adding to his skill and experience from day to day.)

Similar trends are shown by other observers, as shown in Fig. 3.

The results shown in Figs. 2 and 3 were obtained independently of any direct attempt to study the effect of sensitization with onset of experience. Evidence of this kind, together with a great deal of qualitative evidence from previous experiments, and from anecdotal evidence from other work, led to a direct attempt to see if similar tendencies could be shown starting with observers specially selected to have no knowledge at all of the phenomena under investigation.

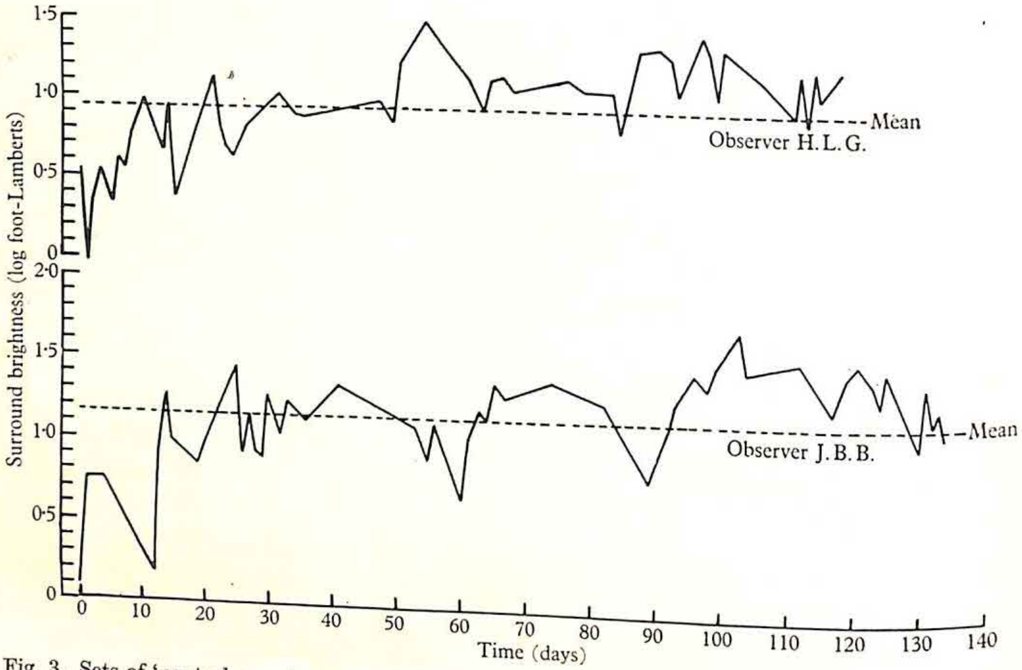


Fig. 3. Sets of 'control experiment' observations given by two further inexperienced observers when subjected to 'just intolerable' glare (source brightness, 3000 ft.L.).

Accordingly two observers were selected for this purpose. These observers were unfamiliar with lighting technology and had never been subjected to any questioning about glare or discomfort in lighting. They were, however, intelligent in their general approach to perceptual problems. Preliminary assessments were made which indicated that they were less sensitive to glare discomfort than the average experienced observer. Each observer subjected himself to the conditions of the control experiment exactly similar to those shown in Figs. 1–3. They were asked to make settings of the environment to give each of the series of degrees of glare discomfort under the controlled conditions, and these assessments were repeated at approximately daily intervals for several weeks. Observations by an experienced observer were made over a similar (but not concurrent) period under the same conditions. The results of this experiment have been reported elsewhere (Hopkinson, 1952) and are reproduced here (Fig. 4). An increase in glare sensitivity similar to that shown by the observer of Fig. 2 is shown here. Whereas the experienced subject has maintained his criterion over the period, the inexperienced subject has gradually crept up in sensitivity (i.e. increased his setting of environmental brightness to balance out glare from the given glare sources) until his settings approached those of the experienced observer. After about 17 days his settings appear to have settled down to a fixed value. The other observer who performed the experiment showed the same trends. Unfortunately, the test could not be continued beyond the point shown in the illustration. Similar curves to those of Fig. 4 were obtained for other settings of the physical environment and for other criteria of glare discomfort. In all, twelve curves of this kind were derived for each observer, all showing the same tendencies.

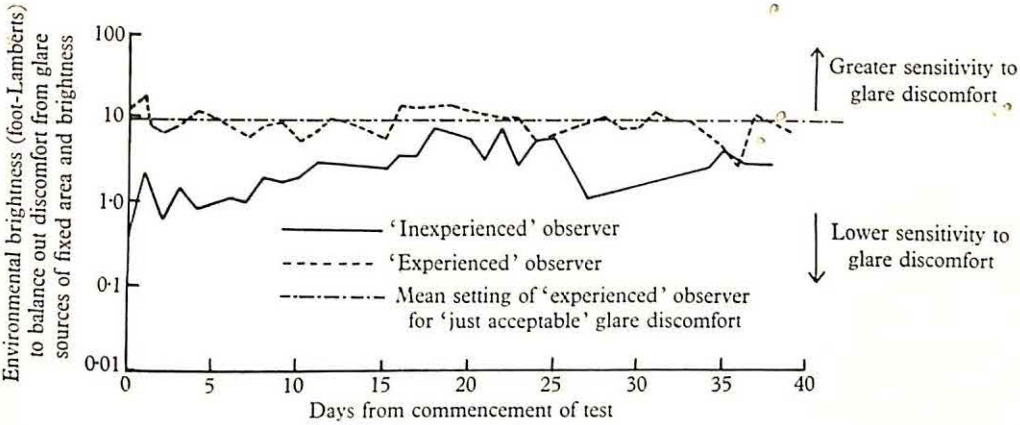


Fig. 4. Glare-discomfort assessments of an 'inexperienced' and an 'experienced' observer under identical conditions over periods of about 40 days for 'just acceptable' criterion, showing increasing sensitivity of 'inexperienced' observer to glare-discomfort over the period.

These observations give quantitative confirmation of the qualitative evidence referred to earlier in the paper. In the specific case of glare discomfort it is well known, from evidence accumulated from a long series of field studies of lighting, that a subject may be unconscious of glare discomfort from a bare lamp over his desk until his attention is directed to it by an investigator. He may then become aware of glare discomfort not only in his working conditions but in other environments as well. He has, in fact, become sensitized to his environment.

IV. STABILITY OF THE OBSERVER

The question remains whether the condition of the experienced observer is one of greater stability than that of the inexperienced observer. Insufficient evidence has been collected to lead to any firm conclusions. All the qualitative evidence does indeed point clearly towards the conclusion that the experienced observer retains his criterion over long periods of time, but the statistical evidence to support this opinion could with advantage be more conclusive. Fig. 5 which represents observations taken under control conditions by three observers over the period from September 1948 to January 1952, is apposite to the discussion. Two of these observers (P.P. and R.G.H.) come into the category of experienced subjects. The third (J.L.) was inexperienced when the observations first started, but subsequently made a great many readings. The observations, which are shown plotted, are again of the environmental brightness necessary to balance out glare discomfort from glare sources in the field of view to yield each of four criteria of glare sensation.

The observations made by the two experienced observers could be taken to support the claim that the experienced observer maintains his criterion.

There is little change in sensitivity to glare discomfort over the very long period under investigation, with some very long breaks between series of observations.

It is indeed remarkable that observers can maintain a criterion of a supposedly vague sensation such as 'discomfort' over such a very long period. Such variations as there are, are less than the differences between one criterion and the next, i.e. between 'just uncomfortable' and 'just acceptable'.

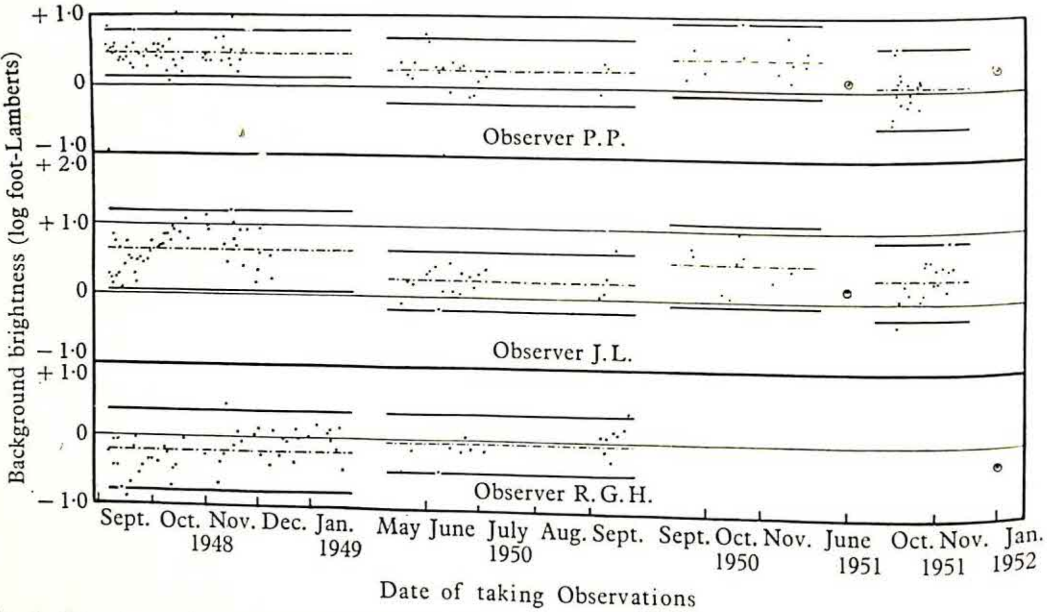


Fig. 5. Long-term glare sensitivities of three observers engaged in discomfort glare control experiments. Each dot represents an observer's setting of background brightness to give 'just uncomfortable' glare when the brightness of the sources is set at 1000 ft.L. The mean value and 95% confidence limits of each group of observations is shown. Single observations represented by '⊙'.

Fig. 5 shows mean values for each batch of observations, and the variance limits (in this case twice the standard deviation) are also shown for each observer's settings. The mean values and variance limits of the same observations have been analysed, and in addition the mean values and variance limits for the three other criteria of glare discomfort (just intolerable, just acceptable and just imperceptible glare discomfort). Similar sets of observations exist for other levels of source brightness.

On the whole, the observations confirm that the condition of the experienced observer is one of fair stability and therefore that such an observer can be used over an extended period of time as a meter of his surroundings. The experienced observers have held their criteria of glare discomfort over a period of three years with little change. If anything there is a tendency to a reduction in glare sensitivity, particularly by observer P.P., but the true significance of this change could only be assessed if observations were available for a much longer period of time.

The third observer, J.L., shows a wider variation in his settings. After his initial 'sensitizing period', he followed by a reduction in glare sensitivity rather in the nature of a slow but steady decrease towards, but not reaching, the lower sensitivity with which he originally started the series of experiments in his inexperienced condition. This observer is more erratic in his day-to-day observations and so this decline in sensitivity has no greater statistical significance than the less marked changes in sensitivity of the first observer.

The rest of the data (for other observers, and other conditions) does not add materially to the picture presented above.

(1) The evidence is definitely in favour of the retention by the already experienced observer of his criterion.

(2) It is also definitely in favour of the existence of a 'sensitization period' for the inexperienced observer.

(3) The subsequent behaviour of the inexperienced observer after the sensitization period is over (he now having become experienced) is not established. Qualitative evidence is in favour of his ability to retain the criterion established at the end of his probationary period, but the quantitative evidence is inconclusive.

V. THE SELECTED TEAM AND THE GENERAL POPULATION

It is hoped to continue the observations for a further period in order to establish whether or not subjective studies should, for preference, be conducted with the aid of a small team of experienced observers, each making a large number of measurements, or whether the preferable technique is to obtain a large number of measurements from single observations from a large number of observers selected at random from the population. From the practical point of view there can be no doubt about the answer. It is much more practicable to work with a small team of experienced observers, who in general make their observations quickly and confidently, and who are available in the laboratory when they are wanted, than it is to attempt to collect a very large number of observers and to make sure that they are selected from the correct cross section of the population. In fact it is almost certain that much subjective study work would have to be abandoned if the latter course were indicated.

The subjective studies in lighting at the Building Research Station are conducted with a small team of selected experienced observers in order to establish the relations between the relevant physical variables. These relations are then tested at one or two significant points by obtaining a large number of observations at these points from a cross-section of the relevant population. This is the only practical course open under present conditions.

The results of such comparisons between the observations of the team and the observations of the population are consistently in the direction of the greater sensitivity to discomfort of the team as compared with the population. The question then arises, how is the information obtained from the team to be used for the specification of conditions to be used by the population as a whole?

The problem at the moment is being solved in the following way. It is possible to prepare, in the form shown in Fig. 6, some of the extensive observations which have been collected from a selection of the general population. This relates the physical variable responsible for the particular sensation of visual discomfort to the proportion of the population who will suffer discomfort at any given level of that variable. The mean value of the physical variable given by the team to correspond to the same criterion of discomfort sensation is also shown in Fig. 6 for the relevant conditions. It is seen that in this case approximately 85 % of the population would experience a lower level of discomfort than the team. Comparable probability curves have been drawn for other situations. It has been found that the mean values of the physical variable which the team observations yield correspond fairly well to a value of about 85 %.

It has therefore been decided for the moment that in recommendations to bodies concerned with the specification or the recommendation of physical levels concerned with visual comfort (for example levels of illumination for visual tasks, levels of brightness for lighting units, etc.) the mean values of the physical variable as given by the team of observers will be offered. There is then evidence that approximately four-fifths of the population would also be satisfied by these conditions. This seems a preferable course to taking the mean value given by the population as a whole, which of course would lead to half only of the observers of the population being properly accommodated. When further information is available on the nature and reasons for these differences between the judgements of experienced and inexperienced observers it may be possible to come to conclusions based on less arbitrary decisions.

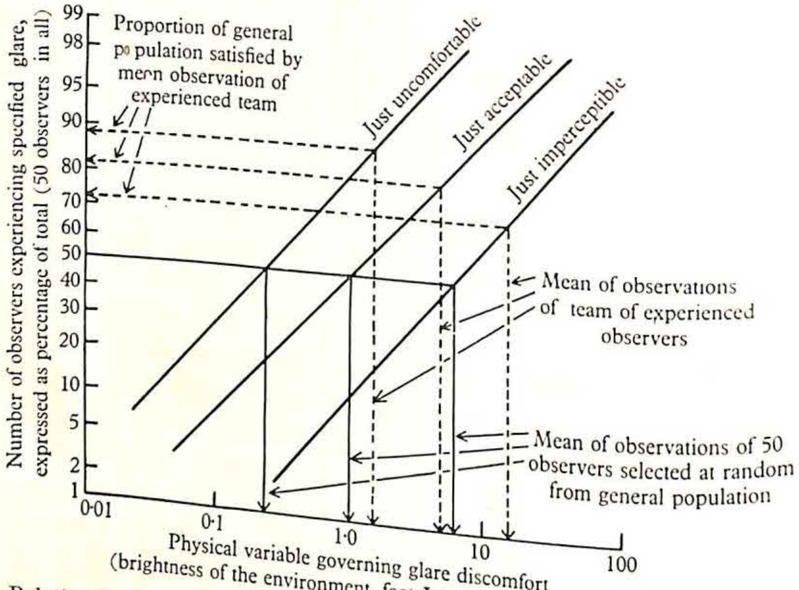


Fig. 6. Relation between mean settings of team of experienced observers and the observations of the general population on glare discomfort.

VI. Discussion

The nature of the study demands a very long time to complete the work. The particular study, that of the estimation of glare discomfort in lighting, is also not one which lends itself readily to a high degree of repetition accuracy. However, it seems reasonable to draw the following conclusions from the observations which are so far available:

(1) In general, inexperienced subjects are less demanding in their requirements to avoid visual discomfort, indicating that they are therefore less sensitive to glare discomfort than experienced subjects.

(2) Some inexperienced subjects when put through an identical series of tests over a long period become sensitized, and their requirements then agree with those of experienced subjects.

(3) Some subjects appear to reach a steady state beyond which no further experience of the situation creates further demands. This is in accord with qualitative findings. There may in fact be a subsequent decline from the state of maximum sensitization. The evidence is clearer for the steady state than for the subsequent decline in sensitivity, but observations over a longer period are needed to establish statistical significance.

Beyond these conclusions it would be dangerous to go at present. The findings reported in the paper will probably confirm in their opinions those workers who already prefer to work with a small team of experienced observers.

There is another factor in the problem which must be brought into the discussion. Most workers in this field will have recognized the effect of personality in subjective judgement. A 'good observer' is co-operative and intelligent, is quick and keen to appreciate the points of technique of the experiment, and so learns to make his judgements readily and rapidly. A good observer also has a social conscience, and gives of his best, without in any way needing to have any knowledge of the trend of his results. It is indeed necessary that the observer should have no preconceived theories about the experiment on which he is engaged. Otherwise what has happily been called his 'coefficient of involuntary mendacity' may bias his judgements for or against his theory. The observer should be indifferent about the *result* of the experiment, but conscientious for the *success* of the experiment. The observer who makes no attempt to master the experiment, and who takes no care to relate his settings to his sensations, in fact who 'couldn't care less', is a nuisance, and his influence can only be eliminated by personal selection of the small team. The timid observer who is over-anxious to give a 'right' judgement is equally a nuisance for a different reason. Ideally such personal factors should be eliminated by random selection of observers, but with the small team technique this is of course not possible.

The practical implications of the conclusions are considerable. They raise economic factors; the experienced observer will demand higher quality equipment (for example for the reproduction of music or of pictures) for the same criterion of excellence as will the 'man in the street'. If the designer decides to plan only for the 'man in the street' he is able to produce a cheaper piece of equipment than if he designed for the experienced observer. This economic factor sometimes works in reverse. An experienced observer can often work usefully with inferior equipment; the novice may be incapable of detecting on a warning display signals which are readily apparent to the experienced man.

In other cases, design for the experienced observer may demand a change in design rather than a more expensive design. In the design of the lighting of buildings the elimination of glare discomfort must be more rigorous if the experienced observer is to be satisfied than if the judgement is by the average man. But here it is a matter of the redistribution of light to illuminate the surroundings as well as the work, to reduce the brightness of light sources, rather than the provision of more expensive lighting equipment.

Finally there are the implications of the conclusions on the methodology of subjective judgements. It is clear that any single judgement of a situation is subject to considerable variance. Even experienced observers deviate at times from their main judgements. What seems of greater importance, however, is the significant differences in the judgement of a situation by experienced and by inexperienced observers. The awareness of discomfort is associated with experience of a situation, but can it be established that no harm comes to the inexperienced subject who is happily unaware of discomfort which is readily appreciable by the experienced man? Much more work would be necessary to establish the true state of affairs.

The existence of an 'experience factor' is of particular importance in view of the increasing use of the questionnaire techniques in social methodology. These are usually based on the single assessments of a large number of inexperienced observers. The error due to the variance of the judgements is eliminated in the statistical procedure of analysis, but the 'experience factor' remains. Can it be confidently accepted that it can be ignored?

VII. SUMMARY

Experimental evidence has been obtained which shows that inexperienced subjects are less susceptible to glare discomfort than those who have had the opportunity to make a long series of judgements. A steady increase in sensitivity has been recorded. On the other hand, experienced subjects retain their enhanced sensitivity at a constant level over a period of years. Such experienced subjects are therefore valuable as human meters of their environment.

The comparison which has been made between judgements of such experienced subjects and those of inexperienced subjects shows that the average judgements of experienced subjects will satisfy about 85% of the inexperienced subjects.

Although all these judgements refer to a study of glare discomfort, it is believed that a similar relation exists in many other situations where judgement of an attribute allied with discomfort is involved.

The work described in this paper forms part of a research programme of the Building Research Board of the Department of Scientific and Industrial Research and the paper is published by permission of the Director of Building Research.

The data recorded in the paper were all taken incidental to an investigation into the physical variables which govern the degree of discomfort in lighting. The author gratefully acknowledges his indebtedness to his colleague Mr P. Petherbridge who conducted the experimental work, and to the observers who submitted themselves to the experiments.

REFERENCES

HOPKINSON, R. G. (1950). The multiple criterion technique of subjective appraisal. *Quart. J. Exp. Psychol.* II, 124–131.

HOPKINSON, R. G. (1952). The influence of experience on the sensitivity to discomfort. *Nature, Lond.*, CLXIX, 40.

PETHERBRIDGE, P. & HOPKINSON, R. G. (1950). Discomfort glare and the lighting of buildings. *Trans. Illum. Engng. Lond. Soc.* XV, 39–79.

(*Manuscript received* 19 *August* 1954)

CHANGES IN TIME PERCEPTION INDUCED BY AN ANAESTHETIC DRUG

By HANNAH STEINBERG

Department of Pharmacology, University College, London

I. *Introduction* (pp. 273–274). II. *Methods and Procedure* (pp. 274–275). III. *Results* (pp. 275–277). IV. *Discussion* (pp. 277–278). V. *Summary* (p. 278). *References* (pp. 278–279).

I. Introduction

Various changes in the perception of time have been reported following administration of drugs such as hashish (Pick, 1919; Bromberg, 1934), mescaline (Fröbes, 1923; Favilli, 1937; McKellar, 1954) and quinine (Sterzinger, 1938). Most of the data published have been observational and qualitative and there seems to have been little systematic follow-up.

During experiments with the anaesthetic gas nitrous oxide carried out by the present writer, some subjects spontaneously commented that time passed extremely slowly, e.g. 'minutes seemed like hours', while others said that time seemed to 'fly'; and a few described the effects as complete loss of sense of time. The purpose of the investigation to be reported was to explore whether such phenomena could be measured quantitatively, i.e. whether there was a relation between subjectively experienced changes in the rate of passage of time and estimates of the duration of time intervals made in accordance with a standardized procedure.

Some previous investigators have suggested that this might be so. Hoagland (1935) has sought to account for the impression of a patient suffering from influenza with fever that 'time seemed to pass very slowly' by assuming that the increased body temperature had made the patient's internal chemical timing mechanism run faster. This, he argued, 'should make time, as judged by objective clocks...appear to go by more slowly, since more psychophysiological "time" would pass in a given constant unit (a minute, let us say), of our conventional constant time scale standard'. In order to check, Hoagland asked the patient to estimate the duration of 60 seconds by counting. She completed the counts in considerably less than a minute, especially when the fever was relatively high. In a recent review Woodrow (1951) has referred to this finding as indicating that 'what was actually a minute by physical time seemed longer than a minute to the subject—an interpretation in agreement with the subject's report that time seemed to pass very slowly'. Case & Haldane (1941) have reported that during inhalation of air with carbon dioxide at high pressure time seemed to some of their subjects to be passing more quickly than usual, and in support they cite the finding that one of them took 60 seconds to tap out what he believed to be 30 seconds, while under normal conditions he took only 28. Assumptions about relations between subjective experience and quantitative estimates of time seem also to underlie conclusions reached by other workers. Petrie (1952) found that after leucotomy subjects tended to reproduce demonstrated time intervals as shorter, and to estimate verbally the length of time spent on a given task as longer than before operation. Petrie interpreted this as suggesting that 'time seems to be passing more quickly after leucotomy'. To illustrate the 'universal experience' of hashish users that 'time appears to be markedly lengthened', Bromberg (1934) has quoted the case of a subject who realized after the period of intoxication that what seemed like several days had in fact been only a few hours. Conversely, Barach & Kagan (1940) have concluded from the fact that under the influence of anoxia time intervals were estimated as shorter than actual, e.g. 3 hours spent under anoxia were subsequently judged to have been 2 or less, that 'the perception of time was slowed'.

These various observations are difficult to interpret because of divergencies in procedure, especially in the methods used for assessing time perception. However, for the case of counting out seconds—one of the commonest techniques and the one adopted in the present investigation—the prevalent assumptions seem to be as follows: If for an individual time seems to be passing subjectively 'more quickly than usual', he will count out seconds more slowly than before, since the interval of clock time which he normally thinks to last one second will now appear to him shorter than previously and he will therefore wait longer before judging one second to have elapsed. Hence his 'corrected' estimate of a second will be objectively longer. On the other hand, if time appears to be passing 'more slowly than usual', the subject will think that time is up when objectively it is not yet up, and seconds will be counted out more quickly than before. The experiments to be reported enabled these assumptions to be tested.

II. METHODS AND PROCEDURE

The experiments were carried out in conjunction with others on the effects of nitrous oxide on various aspects of cognitive behaviour (Steinberg, 1954) and were fitted into a common design and procedure. A simple method of difference design was used involving two groups of 50 subjects each, matched for age and sex, and two experimental trials separated by 1 week (Table 1). In trial I treatment was identical for both groups, and subjects carried out tasks under ordinary conditions. In trial II the drug was administered to experimental subjects through a face mask connected by rubber tubing to gas-flow meters. The concentration was 30% nitrous oxide in oxygen. Control subjects received similarly scented air from the same apparatus.

Table 1. *Experimental design*

Group	Treatment	
	Trial I	Trial II
Experimental (50 subjects)	Estimation of seconds (Cognitive tasks) Estimation of seconds: re-test	DRUG Estimation of seconds (Cognitive tasks) Introspections Questions
Control (50 subjects)	Estimation of seconds (Cognitive tasks) Estimation of seconds: re-test	NO DRUG Estimation of seconds (Cognitive tasks) Introspections Questions

Time perception in both the experimental and control groups was investigated by the following methods:

(i) *Estimation of seconds*. The subject was required to estimate the duration of 30 seconds by counting aloud at an estimated speed of one count per second. The time he actually took to complete the count of thirty was measured. In trial I, subjects carried out various cognitive tasks for half an hour after completion of this test and were then re-tested, so that measures of reliability could be calculated. No re-test was given in trial II.

(ii) *Subjective assessments.* On completion of the cognitive tasks in trial II, gas administration was stopped and the subject was asked to write down full introspections on any experiences he might have had during inhalation. This gave him an opportunity to comment spontaneously on changes in time perception. After the introspections had been written, the experimenter asked a number of questions about the effects of inhalation, one of which was whether the subject had noticed anything unusual about the passage of time during the experiment. If he answered in the affirmative, he was asked to specify the nature of the change.

III. RESULTS

(1) *Subjective assessments*

Subjects' spontaneous introspections were searched for references to time perception, and both they and the answers to the experimenter's subsequent questions were classified in three categories, according to whether time during inhalation had appeared to pass (i) more quickly than usual, (ii) more slowly than usual, or (iii) at a fluctuating or unspecifiable rate. This analysis was carried out separately for the spontaneous introspective material and for the answers to the experimenter's questions. The results are summarized in Table 2.

Table 2. *Subjective assessments*

	Experimental group		Control group	
Report on passage of time	Spontaneous introspections	In answer to questions	Spontaneous introspections	In answer to questions
'Quicker'	13	26*	0	12
'Slower'	4	14*	0	10
Miscellaneous	2	7	0	0
No change reported	31	2	50	14
No. of subjects	50	49†	50	36†

* One subject is included whose replies to questions contradicted his spontaneous introspections.
† For practical reasons it was not possible to ask all subjects the questions.

It will be seen that nearly all experimental subjects reported some change in time perception, though only 19 of them did so spontaneously. In their spontaneous introspections 13 experimental subjects reported 'quicker' passage of time, and altogether 26 said 'quicker' in reply to the experimenter's questions; 12 of these had already done so in their introspections, but one subject contradicted his introspections in answering the questions. Four subjects reported 'slower' passage of time in their introspections, and 14 made this report in answer to the questions. Again one subject gave self-contradictory reports. Few subjects gave answers appropriate to the third category of classification. In the control group no subject reported any time changes spontaneously, though in answer to the questions 12 said 'quicker' and 10 'slower'. Since the experimenter's questions produced more responses from the experimental group than had been given spontaneously, and elicited responses from the control group where none had been given spontaneously, it seems possible that subjects were to some extent suggestible. Answers to the experimenter's questions may therefore be less trustworthy than subject's spontaneous reports.

(2) *Estimation of seconds*

The results for the two groups of subjects for each of the experimental trials are summarized in Table 3.

(i) All mean values, including those for measures scored under normal conditions in trial I, are greater than 30, i.e. there was a tendency for seconds to be counted out more slowly than clock time. Individual variation was considerable, however, as is shown by the size of the standard deviations and ranges.

Table 3. *Estimation of seconds: summary of results*

Measures in seconds	Experimental group ($N = 50$)			Control group ($N = 50$)			Diff. between diffs.
	Trial I	Trial II	Diff.	Trial I	Trial II	Diff.	
Mean	34·98	40·68	+5·70	33·20	38·00	+4·80	−0·90
S.D.	9·06	10·23	—	9·17	10·62	—	—
Range	19–70	23–65	—	13–63	13–71	—	—

(ii) The two groups of subjects did not differ significantly in respect of initial performance in trial I ($t < 1$).

(iii) In trial II the means are significantly greater in both groups, as compared with trial I ($t = 4·27$ and 4·66, D.F. = 49, $P < 0·001$). The change of performance in the experimental group from trial I to trial II did not differ significantly from that in the control group ($t < 1$). The drug therefore cannot be regarded as having had any effect on estimating seconds.

That the mean scores of both groups should have risen at all may possibly be an artefact attributable to the apparatus used to administer the gases in trial II. Several subjects reported that wearing the mask made their breathing more deliberate and their natural rhythm of counting aloud was sometimes impeded since they could only talk 'in between breaths'. The possibility that some such external factor was responsible is strengthened by the relatively high reliability of the test. Product-moment correlations between initial scores and re-test scores in trial I were found to be 0·74 for the experimental and 0·89 for the control group. The correlations between scores in trials I and II were 0·61 and 0·69 respectively. The smaller size of the latter pair of correlations may be due to the longer time interval which elapsed between trials I and II—one week as against half an hour—and to individual differences in adjustment to the apparatus and the effects of inhalation.

(3) *Comparison between estimation of seconds and subjective assessments*

The finding that the experimental group and the control group do not differ so far as estimating seconds is concerned does not automatically rule out the possibility of a consistent relation between estimations and subjective assessments. In accordance with the assumptions set out in the introduction, individuals subjectively 'slower' during trial II should have counted out seconds more quickly than before, and vice versa. In order to check on this possibility, the differences in estimation times between trials I and II were calculated for the experimental subjects. The mean value of these differences for the 26 subjects who had replied 'quicker' to the experimenter's questions was found to be almost identical with that of the 14 who had responded 'slower', as is shown in Table 4. A similar result was found to hold for the control group. When the comparison was confined to the 17 experimental subjects who had made spontaneous reports of 'quicker' or 'slower', the result was again similar.

It may be concluded, therefore, that in the present experiments no relation existed between changes in estimating the length of seconds and subjective impressions of changes in the rate of passage of time.

Table 4. *Mean differences in estimation times between trials I and II, in relation to answers given to experimenter's questions*

	Experimental group		Control group	
Answers to questions	N	Mean difference in estimation times between trials I and II (seconds)	N	Mean difference in estimation times between trials I and II (seconds)
'Quicker'	26	5·46	12	8·10
'Slower'	14	5·50	10	3·75

The difference between the 'quicker' and 'slower' subjects is not statistically significant in either the experimental or the control group.

IV. DISCUSSION

In the light of the results described, the original assumptions cited in the introduction were re-examined. Since it had been shown that changes in subjective experience were not reflected by changes in counting out seconds, it followed that the latter did not constitute an appropriate measuring device for the former. The following reasons are suggested:

First, the assumption that subjective impressions of quicker passage of time should lead to slower counting of seconds and vice versa is logically not compelling, since an equally plausible case may be made out for the opposite relation: thus, if for an individual time appeared to be passing 'more quickly than usual', he could also be expected to count seconds more *quickly*, since it would appear to him that the moment for counting the next second was already there when according to clock time it had not yet arrived. The apparent paradox is probably attributable to the ambiguity of statements like 'time passed more quickly (or more slowly) than usual'; the reference may, for example, be primarily to some sort of 'internal clock', the speed of which is somehow perceived to have altered, or it may be to the rate at which events in the outer environment are experienced. Such a distinction is illustrated by comments like the following: 'Overall time seemed longer—it's hard to describe—I had more time to do the tasks, but time itself was passing more quickly.'

Secondly, one could not be sure in the present investigation of the extent to which estimates of time may have been influenced by objective frames of reference, and of how far subjects may have tried to make corrections. Suppose, for instance, that during the drug administration time seemed to a subject to be passing more slowly, in the sense that he seemed to have more time for a certain task than on a previous occasion, but other cues led him to guess that the objective time must be the same: there then is no reason why he should deliberately try to compensate by altering his usual speed of counting out seconds. He could (but equally well might fail to) realize that although his usual speed now appeared to him, say, slower, it was objectively probably still the same as before and therefore, presumably, correct.

Thirdly, two different kinds of time perception seem to be involved. Counting out seconds requires judgements of the duration of time intervals; the subjective impressions, on the other hand, refer not so much to duration as to rate of passage of time. Conceivably, statements about the former can be made independently of the latter. For example, a time interval may in retrospect be judged to have appeared shorter than its objective duration, yet while it was in progress time may have been experienced as passing 'more quickly' or 'more slowly' or 'at the same rate as usual'.

From the above arguments it follows that any given judgement of time, whether impressionistic or 'objective', may be determined by a complicated interplay of subjective and contextual factors; and there are no theoretical or, as the results of the present investigation have confirmed, consistent empirical grounds for expecting impressions of changes in the rate of passage of time and estimates of durations of time intervals to be systematically related. Nevertheless, investigations like those quoted earlier, in which a relation of this kind was apparently observed, seem worth pursuing further and with larger numbers of subjects and the co-ordinated use of different methods of assessing time perception. It may become possible to identify quite specific aspects of time perception which are selectively susceptible to drugs and other physiological influences. Apart from their intrinsic interest, such findings could be utilized for discriminating between various factors involved in time perception and for studying their respective roles.

V. SUMMARY

Effects of an anaesthetic drug (nitrous oxide) on subjective impressions of changes in the rate of passage of time were studied, and were compared with performance in a test of estimating the duration of seconds by counting aloud. The assumption was examined that subjective impressions of 'quicker' passage of time should be reflected in slower counting of seconds and vice versa. The results showed that:

(1) The drug induced subjective impressions of changes in the rate at which time was passing in a substantial proportion of subjects, some of whom reported apparent quickening and a smaller number apparent slowing.

(2) The drug had no effect on estimating the duration of seconds by counting.

(3) No relation could be demonstrated between subjective impressions of changes and changes of performance in the test of estimating seconds.

Possible reasons for these findings are discussed.

I am grateful to Prof. F. R. Winton for his help throughout this investigation; to Prof. R. W. Russell for advice and encouragement; to Mr A. Summerfield for helpful discussion; and to the Medical Research Council for financial assistance.

REFERENCES

BARACH, A. L. & KAGAN, J. (1940). Disorders of mental functioning produced by varying the oxygen tension of the atmosphere. *Psychosomatic Med.* II, 53–67.

BROMBERG, W. (1934). Marihuana intoxication. *Amer. J. Psychiat.* XIV, 301–30.

CASE, E. M. & HALDANE, J. B. S. (1941). Human physiology under high pressure. I. Effects of nitrogen, carbon dioxide, and cold. *J. Hyg., Camb.*, XLI, 225–49.

FAVILLI, M. (1937). La percezione del tempo nell' ebbrezza mescalinica. *Rass. Studi psichiat.* XXVI, 455–62.

FRÖBES, J. (1923). *Lehrbuch der experimentellen Psychologie.* Freiburg: Herder.

HOAGLAND, H. (1935). *Pacemakers in Relation to Aspects of Behaviour.* New York: Macmillan.

MCKELLAR, P. (1954). Personal communication.

PETRIE, A. (1952). *Personality and the Frontal Lobes.* London: Routledge and Kegan Paul.

PICK, A. (1919). Psychopathologie des Zeitsinns. *Z. Pathopsych.* III, 430–41.

STEINBERG, H. (1954). Selective effects of an anaesthetic drug on cognitive behaviour. *Quart. J. exp. Psychol.* VI, 170–80.

STERZINGER, O. (1938). Neue chemopsychologische Untersuchungen über den menschlichen Zeitsinn. *Z. Psychol.* CXLIII, 391–406.

WOODROW, H. (1951). In Stevens, S. S. *Handbook of Experimental Psychology.* New York: John Wiley and Sons.

(*Manuscript received 8 October 1954*)

SOME FACTORS INFLUENCING THE EFFECTIVENESS OF AN INSTRUCTIONAL FILM

By S. LANER

The Department of Psychology, University of Reading

I. *Introduction* (p. 280). II. *Method* (pp. 280–282). III. *Results* (pp. 282–285): (i) *Analysis of sketches* (p. 282); (ii) *Analysis of responses to oral questions* (pp. 282–283); (iii) *Comprehension and assembly performance* (pp. 283–284); (iv) *The factor of past experience* (p. 284); (v) *Reports of film-instructed subjects* (pp. 284–285). IV. *Conclusions and Discussion* (pp. 285–288). V. *Summary* (p. 289). *Appendix* (pp. 289–292). *References* (p. 292).

I. Introduction

The rapid strides in the development of pictorial devices for communicating information and instruction has led to the hope that the moving film and television might come to supplant rather than supplement predominantly verbal methods of teaching and training. This hope appears to be based on the belief that pictorial material makes a greater impact on the learner which is liable to increase in proportion as all features of objects, situations and events, particularly motion, are faithfully represented.

But there is growing evidence that increased fidelity of representation is not *per se* an effective substitute for clarity and soundness of exposition. In a previous experiment (Laner, 1954) no significant differences were found between the performances of subjects on a manipulative task after instruction by either a highly realistic moving film or by a film-strip composed of line drawings, the pictorial material in the latter being obviously more in the nature of supplementary illustrations. In particular no special effects were found that could be attributed to 'realism' or 'motion'.

The experiment described in this paper attempts to follow up and extend some of these findings. It was designed so as to bring out as clearly as possible whatever special effects might be contributed by realistic portrayal, especially represented motion. Towards this end: (*a*) a moving film was selected showing the operation of a relatively complicated and unfamiliar piece of mechanism, which could only be understood if the interactions of its various parts in motion were appreciated; (*b*) this film was pitted against a corresponding text, supplemented merely by two static diagrams. The subjects, having been instructed by one of the two devices, were required to sketch and explain the mechanism, as well as design their own procedure for assembling it.

It was further attempted to analyse out factors in film communication which appear to influence the perceptual process in specific ways, and thereby contribute to or detract from the instructional effectiveness of a film; to trace such factors and define their effects is among the chief aims of the research of which the present study forms a part.

II. Method

The display material

The two types of display material used contained an identical account of the mode of operation of the Bren-gun trigger mechanism. The main considerations which led to the selection of this content were the great emphasis given to represented motion in the film, as well as the fact that even subjects who had previously handled a Bren gun could be relied on to be unfamiliar with the trigger mechanism, the dismantling of which is normally an exclusive prerogative of trained armourers.

The film passage, shown individually to one group of subjects, forms part of a complete explanatory film of the Bren gun, produced by Publicity Productions Ltd. for the instruction of Army armourers. During the experimental showings it was always preceded by an introductory passage, which shows the general mode of operation of the whole gun and which provided the necessary context. The whole showing lasted 10 minutes 15 seconds.

The text used to instruct a second group of subjects was also administered individually, and was a literal transcript of the film commentary, except that references to the display and repetitive statements were deleted. A general description of the way in which the Bren works was included in it. Two static diagrams were appended. (For text and diagrams, see Appendix.)

Subjects

The experimental population consisted of fifty male subjects, all University students of whom half saw the film; the other half studied the text. None of the subjects had had any previous experience with the trigger mechanism.

Procedure

As a preliminary to each session all subjects were told of the purpose of the experiment. It was pointed out to them that the initial passages in the film and the text were merely orientational, but that comprehension of this passage would provide them with the necessary context for understanding the trigger mechanism and should therefore be attended to. The text-instructed subjects were in addition told that they would be allowed 10 minutes 15 seconds to study the text and that they were free to make any use they wished of the diagrams.

Testing procedures

Immediately after viewing the display each subject was asked to draw a rough sketch of the trigger mechanism and to insert any names of parts that he could recall. The manner in which the display had been utilized was then inquired into. Next, comprehension was tested by a set of carefully phrased questions. This oral inquiry made the situation less constrained and made it possible to demand further clarification, where the initial explanation was not deemed sufficient. Special care was taken not to give any hints. Eighteen questions were asked, e.g.: 'What is the effect of the Change Lever on the Tripping Lever?', or 'What is the function of the double claw with the gun set at repetition?'

Finally the subject was given the seven component parts of the trigger mechanism and asked to assemble them. The manner of assembly was not shown in the film or described in the text, and the subject had to evolve it for himself. The time taken to complete the assembly was recorded.

Scoring and analysis procedures

The three sets of data obtained were analysed as follows:

(i) The sketches were classified into two categories according to whether the mechanism as drawn could be held to 'work' or 'not work'. To eliminate such irrelevant features as, for example, artistic merit, the sketches were analysed part by part and points were awarded for (*a*) inclusion of a part, (*b*) 'functional' shape of the part (would it work as drawn?), and (*c*) 'functional' position of part. It was found that drawings with a score of 9 points or less out of 15 did not meet the criterion of 'workability'.

(ii) The subject's responses to the oral questioning fell readily into four categories: correct answer, doubtful, wrong, no answer. A comprehension score for each subject was obtained by summing his correct replies and converting into percentages.

A comparison of the total percentage scores of the subjects in the two groups, however, allows only a comparison of 'overall effectiveness'. By summing the responses of each whole group of subjects to each individual question, it was possible to obtain, in addition, an indication of display effectiveness at each stage. The replies were summed into a 'Right' and 'Wrong' category for each question, the 'Right' category comprising all correct responses, the 'Wrong' category all incorrect responses as well as failures to respond (responses classified as doubtful were left aside). The resulting 2 × 2 tables were treated by the χ^2 technique.

The times taken by subjects to assemble the trigger mechanism provided scores which could be treated in a statistically straightforward manner.

III. RESULTS

(i) *Analysis of sketches*

The pattern into which the workable and unworkable sketches distributed themselves is shown in Table 1.

Table 1

	Workable	Not workable
Film group	10	15
Text group	8	17

The expectation that the film group would do better on the sketches was not supported by the χ^2 value of 0·34, which does not even approach significance. It should be added that the subjects did not know beforehand that they would be asked to draw a sketch. Their performances tended to be halting and, as far as one could judge, looked like attempts at construction. It is, furthermore, noteworthy that most subjects tended to recall the name of any given part first before drawing it.

Table 2

	Mean % correct replies	Standard deviation
Film group	63·99	29·05
Text group	53·81	34·38

(ii) *Analysis of responses to oral questions*

As described in the previous section a comprehension score for each subject was obtained by summing his correct replies and converting into percentages. The mean percentage scores and the standard deviations for the two groups of subjects are shown in Table 2.

The approximately 10% difference between the means is not statistically significant ($t=1·11$); moreover, as will be shown, the size of this difference is exaggerated.

The overall comprehension scores, on which the above calculations were based, are not by themselves good indicators of display effectiveness. Under certain circumstances—to quote an extreme example—the film might be more effective in conveying the first five facts or concepts and the alternative method in conveying the final five; identical overall scores would in such a case cover up completely the differential effectiveness of the two forms of instruction. By comparing the response of each group as a whole at each stage this differential effectiveness is thrown into relief. Table 3 shows the χ^2 values obtained by the complementary method.

The χ^2 values for questions 16–18 are significant at the conventional levels, but this significance is purely statistical. They are accounted for by the fact that many of the text-instructed subjects found the time too short to get to the end of the description; answers to the last three questions from these subjects were therefore either inadequate, or (for the most part) absent. It is easy to see that this same fact was also reflected in the text-instructed subjects' overall comprehension scores and through these scores in the mean value, thus contributing to the 10% difference between the mean comprehension values.

Table 3

Question	χ^2	Question	χ^2
1	1·44	10	0·37
2	1·44	11	0·33
3	1·44	12	1·71
4	Not calculable	13	3·12
5	1·00	14	0·03
6	1·33	15	1·56
7	0·95	16	7·59
8	0·52	17	7·14
9	1·28	18	9·62

Of the remaining χ^2 values that for question 13 approaches significance, the advantage here being decidedly with the text-instructed subjects. The question involved the action of a spring, which apparently was not perceived at all by many film-instructed subjects, who also missed its functional importance. The spring was omitted from many of the sketches of the film-instructed subjects, while all the text-instructed subjects included it.

(iii) *Comprehension and assembly performance*

Despite the fact that neither film nor text gave any description of how to assemble the trigger mechanism, all subjects succeeded in the task of evolving an assembly procedure. The time taken to assemble varied widely among individual subjects, although these variations had, apparently, little to do with the nature of the instructional device used: the difference between the mean assembly time of the film- and text-instructed groups was only 1·36 minutes and not statistically significant. It was, however, noted that the time taken to complete the assembly tended to reflect two kinds of approaches: a subject who tackled the task in a purely trial and error fashion usually took a good deal longer than one who chose a less random approach. This in turn made it not unreasonable to test the hypothesis, suggested by a previous experiment, that a more adequate assembly performance and better comprehension are connected.

For various reasons the procedure of correlating comprehension and time scores was rejected; instead the two groups of subjects were each divided into two classes, one comprising all subjects who showed above average understanding on the oral questionnaire, the other comprising all subjects who fell below average. The mean time scores of the two classes in the film and text groups were then compared statistically. The statistics obtained are shown in Table 4.

Table 4

	Mean for 'above average' comprehension (minutes)	Mean for 'below average' comprehension (minutes)	Diff.	t
Film group	5·38	10·55	5·17	3·82
Text group	4·13	9·03	4·90	3·72

The 't' values are significant at the 0·001 and 0·01 levels respectively.

From this table it is quite clear that, regardless of the type of instructional material used, subjects who showed more comprehension of the functioning of the mechanism were also considerably quicker (i.e. better) at constructing for themselves an assembly procedure. Although it would seem to follow that a more adequate performance is based on better comprehension, the alternative hypothesis has to be considered, i.e. that a common factor underlies both comprehension and assembly performance.

(iv) *The factor of past experience*

The term 'past experience' in this context does not denote that the subject already knew how the Bren-gun trigger mechanism works or that he had assembled it on a previous occasion; it must be emphasized that none of the subjects had had such specific experience. A subject was classified as 'experienced' on the basis of his general familiarity with things mechanical, the extent of which was determined by inquiry before each session.

Table 5 shows how the experienced and inexperienced subjects in both test populations compared with one another on the comprehension scores.

There were also differences between the experienced and inexperienced in both groups in the times taken to perform the assembly operation. However, the great variability of the inexperienced subjects precluded a valid comparison of the means.

Table 5

	Mean for experienced (%)	Mean for inexperienced (%)	Diff.	t
Film group	73·00 ($N=14$)	52·52 ($N=11$)	20·48	2·46
Text group	65·65 ($N=11$)	48·80 ($N=14$)	16·85	2·19

Both 't' values are significant at the 0·05 level.

(v) *Reports of film-instructed subjects*

The analysis of qualitative data was in the main confined to reports of subjects in the film-instructed group. Particular interest attaches to those aspects of these reports, which cast light on the stage when the display was actually presented, i.e. when it is making its direct impact. Subjects' reports are, in fact, the only direct means of access to events during the viewing stage, of which the quantitative indices of subjects' performances give at best indirect indications.

In general, the subjects' attitude to the effects and effectiveness of pictorial media of instruction was predominantly positive; but this seems merely to reflect conventional views regarding the merits of predominantly pictorial as against verbal methods of conveying information and instruction. This positive attitude was, incidentally, also shared by most subjects in the text-instructed group.

When asked to substantiate the general view with reference to the film they had just been shown, some subjects suggested that the direct pictorial representation of certain spatial relationships and of the directions of movements of certain parts would give the film advantages denied to, for example, a written text. On the other hand, very few subjects found that pictorial presentation favoured the later re-evocation of specific visual images; in fact, many subjects recognized that what imagery they had had was itself of a constructive rather than reproductive nature; several subjects, for example, stated that they had to substitute their own diagrammatic image-versions for the 'confusing' images supplied by the display.

For the most part, however, the subjects were preoccupied in their reports with the effects in perception of the pictorial information being presented in a continuous series—a characteristic of film and television presentation. A survey of subjects' statements suggested that this type of presentation resulted in the comparatively frequent occurrence of a number of checks and stresses in perception:

(*a*) The meaning or reference of some part of the display was misinterpreted and while an attempt was made to correct for this misperception, further stimuli from the screen could not be dealt with. ('I could not follow at some points and there was a sort of breakdown, due to my interpretation of the film content.')

(*b*) A lag in the perceptual response was brought about by the subject's attempt to reorient himself to a spatial or ideational shift in the display. ('Too rapid as the diagram was twisted round. If you hadn't picked up something, you didn't have a chance to pick it up again.')

(*c*) A temporary fluctuation in attention resulted in what subjects described as a 'loss of the thread'.

(*d*) The perceptual response was checked by some vividly represented or unfamiliar item in the display, which temporarily monopolized the subject's attention. ('Not knowing how "sere" is spelt creates a short distraction and thus cuts across concentration on the film.')

(*e*) The perceptual process was crossed by some motivational influence, e.g. anxiety. ('You develop a lot of anxiety as the thing is racing ahead and this anxiety has a disruptive effect. All the time you are trying to hold on to what went before and missing what is just going on.')

Each of these checks—and the list is probably not exhaustive—resulted, according to the subjects, in a more or less extensive loss of information, which was in many instances cumulative.

IV. CONCLUSIONS AND DISCUSSION

Our present results agree in a number of respects with those of a previous study (Laner 1954) and it is significant that in neither case could any evidence be found for any special effects attributable to the realistic and detailed character of the pictorial representation—such, for instance, as the portrayal of movement. The first study dealt with the execution of a manipulative task after instruction by film or film-strip. In the present experiment the emphasis was upon the comprehension of a mechanism of mutually interacting moving parts. It would seem that in both of these diverse connexions, the realistic representation of the movements involved was relatively unimportant. Our observations indicated that it could even be confusing.

On the other hand, it seems important for clear perception and understanding that the effects or end-results produced by each consecutive movement or sequence of movements and their relationships to the preceding stages of events or operations should be adequately demonstrated. If these relationships are clear to the subject, there seems to be no difficulty in constructing an adequate version of intervening events, whether in terms of actual movements or in terms of verbal explanation. The fact that in the present experiment those subjects who had shown the best understanding were also quicker and better at devising their own assembly procedure supports this contention.

The dismissal of 'realism' as comparatively unimportant, however, has implications which have a bearing on the still highly controversial issue of pictorial *versus* non-pictorial, and visual *versus* auditory presentation of information. It is widely held that realism is the very factor which makes pictorial visual presentation more efficacious; underlying this belief are two assumptions: first, that the eye can take in more information than the ear; and second, that since a picture presents simultaneously a good deal of information, the superior 'carrying' capacity of the eye can be brought fully into play. But, as many studies have shown, the important factor in perception is not the high carrying capacity of the visual channel, but the very much smaller capacity of central mechanisms in handling the incoming information. Moreover, as experiments with brief exposures of pictorial material have shown (e.g. Bartlett, 1932), very little accurate information can be obtained 'at a glance'. It is significant that studies with films, designed specifically to establish the superiority of visual presentation, have produced negative or at best equivocal results (Nelson, 1949; Nelson & Moll, 1950; Yale Motion Picture Research, 1947; Van der Meer, 1950).

It does not, of course, follow that for certain types of information visual presentation may not have advantages over other types of presentation. It does follow that increased efficiency is not an automatic by-product of realistic, pictorial-visual presentation, but that efficacious pictorial communication is subject to a number of conditions, just as are other forms of communication. These conditions are inherent both in the receiving organism and in the display. In the discussion that follows an attempt is made to elucidate some of these conditions with particular reference to continuous serial displays (film, television). Though the arguments advanced are, at this stage of the research, based in the main on subjects' reports, the fact that their evidence is at many points in good agreement with the findings of research undertaken elsewhere renders the account given somewhat less tentative than it would be otherwise.

Of the conditions contributed by the viewer—these have, not inappropriately, been termed 'audience variables' by the Pennsylvania film research team—the most important in the present and previous studies appears to have been related past experience. Apart from the quantitative results obtained to this effect, there are numerous reports by subjects which indicate that past experience not only made it possible to integrate the information actually present, but provided a framework of future reference, which in turn allowed anticipation. That the possibility of anticipation is a material help in perception has been stressed by many subjects.

Our main concern, however, was not with 'audience variables' so much as with factors inherent in the structure of the display. Judging by the frequency of subjects' complaints and remarks, one of the main factors in this category is the so-called rate of development, i.e. the rate at which the information is conveyed. In practice—and this is where 'realism' comes in once more—this rate is commonly governed by the rate at which the represented events actually occurred; occasionally use is made of slow motion. If, however, visual aids are regarded not as devices for copying reality, but as means for effectively conveying information, it at once becomes clear that what must primarily be taken into account is the rate at which any given information content can be dealt with by central processes. In general this rate for any type of audience lags well behind the actual occurrence of events; representation at actual speed then leads to disruption of the perceptual processes by motivational factors, such as anxiety or apathy resulting from frustrated efforts to perceive what is going on. Occasionally, on the other hand, when slowly occurring events are represented at their actual speed, the perceptual process is disrupted by boredom and tedium. The problem is one of determining empirically broad optimal ranges of rate of development for various types of information. Experimental studies along these lines have already been started (Jaspen, 1948, 1950).

Closely akin to the rate of development is the factor of repetition, which has also frequently been mentioned by subjects. Although it can be considered established that repetition as a whole or of certain sequences tends to increase the amount of material assimilated from the pictorial medium, there still remains the problem of the form repetition should take. In this connexion the information theory principle that the probability of getting information across may be only slightly increased by literal repetition has been strikingly paralleled in Jaspen's (1948) study. It was found that the effectiveness of a film 'key' sequence increased in step with the inclusion of one, two, three and four *non-identical* repetitions. Certain indications were found in the present experiment which suggested reasons why literal repetition may be less effective, the chief being probably the interference of boredom and straying of attention.

A further factor which has been little noted so far, but which appears to influence the intelligibility of the visual message to a considerable extent, is the frequency of shifts. By 'shift' is here undersood any sudden extensive change in the display. A spatial shift, for example, is introduced when a long-shot sequence, giving a more inclusive view, gives way to a close-up sequence, where a small part-aspect of the overall view takes up the major proportion of the total screen area; a similar type of shift is introduced by a change in camera angle. Every change of this kind appears to demand some reorientation on the viewer's part. But in addition the viewer must, presumably, also develop some sort of spatial scheme, corresponding to the space in which the camera was moved. And finally he must adapt to sudden changes in sizes. It seems evident that the more frequently such shifts occur, the smaller the chances of a spatial scheme being formed and the greater the likelihood of disorientation and loss of information. Difficulties of this kind were frequently mentioned by subjects, directly or indirectly. Similar difficulties were reported in connexion with a shift in depicted content—an 'ideational' shift—e.g. at the point where the film switches from a representation and description of the Bren gun as a whole to the representation of the trigger mechanism. Here the ideational and spatial shifts occurred simultaneously.

But by far the most important factor influencing the intelligibility and thereby the effectiveness of the visual display appeared to be the spoken commentary. Although the use of many unfamiliar technical terms to accompany an equally unfamiliar pictorial content tended to throw into particular relief the more negative effects of the commentary, yet at the same time it emphasized the extent to which the commentary can affect the perception of the visual element. Numerous remarks made by subjects suggested, in the first place, that as a consequence of their inability to get hold of a technical term they were left to face a relatively undifferentiated welter of visual impressions. ('If you knew names of parts, you would know what they looked like and would be able to follow better.') There was no dearth of reports indicating that the commentary, and particularly certain of its key sequences and terms, can materially assist visual perception by contributing to the structuring and articulation of the perceptual field. ('Knowing names would help in organizing'; 'If I had known the names of parts, I would have been less bewildered at the start'.)

Even more importantly, the commentary seemed vital in communicating functional relationships. While a pictorial display may be adequate in demonstrating certain spatial and directional relationships, more abstract functional and cause-and-effect relationships seem to require explicit formulation in verbal terms before they can be grasped. This is possibly due to the fact that such relationships demand the perceptual bridging of time-gaps as well as the narrow selection of relevant aspects from a display where relevant and irrelevant aspects carry equal weight. In the present experiment it was often the *name* of some part which contained the characterization of a functional relationship (e.g. tripping lever). When, then, for one reason or another, the name was not perceived or retained, the functional relationship remained equally unperceived ('...missing the name one would miss the function of the part, too'; 'Without remembering the names it was impossible to follow...').

Altogether, the evidence converges on the principle that the effects produced by the auditory and visual elements of a combined display—particularly a continuous serial display—are not independent and simply additive. Thus a commentary that overlaps highly in content and formulation with the pictorially presented content can and does materially facilitate pictorial perception. On the other hand, a commentary that has too high a word density (Jaspen, 1948; Zuckerman, 1949), is inadequately synchronized or badly formulated, uses a high percentage of unfamiliar expressions or diverges from the visual element in content is not just ineffective; it tends to reduce the effectiveness of the pictorial element. This is perhaps because instead of smoothing the assimilation of pictorial information it imposes an additional load on the perceptual mechanisms and thus makes the task of handling the incoming information more difficult.

To conclude, it has become increasingly evident to the experimenter in the course of the analysis that from the point of view of assimilation of information the greatest importance attaches to the way in which the message is organized. The problem of pictorial message formulation seems to be, in fact, akin to the problem of 'encoding' which information theory recognizes as being an important consideration with reference to the efficiency with which a message can be handled by the receiver. But since, in practice, visual aid communication almost always involves the simultaneous employment of pictures and verbal and other symbols, and the simultaneous employment of the auditory and visual channels, the problem of pictorial formulation is part of the wider problem of concurrent pictorial-verbal-symbolic formulation.

V. Summary

1. Experiments were carried out in which twenty-five subjects were individually shown a sound film of how the trigger mechanism of the Bren gun works. A comparable group of subjects was given individual instruction by means of a printed text and two static diagrams. Subjects were asked to draw a sketch of the mechanism, explain how it operates and assemble a model.

2. In terms of overall scores no significant difference was found that could be attributed to the effect of either instructional device. Furthermore, neither device was more effective in conveying any part of the instructional content.

3. A significant positive relationship was found between comprehension scores and the time taken to evolve and carry out an assembling procedure.

4. A significant positive relationship was also found between adequacy of explanation and the amount of related past experience.

5. It was concluded on the basis of the quantitative, as well as qualitative, data that realistic pictorial portrayal contributed little to instructional effectiveness. An analysis of subjects' reports suggested a number of factors in the film medium which appear to be more important in this respect.

The continued financial support given this research by the Trustees of the Nuffield Foundation is gratefully acknowledged. Dr M. D. Vernon has readily offered her advice at both the experimental and report stages of this investigation. To Prof. R. C. Oldfield I am indebted for his steadfast sponsorship and interest in the present experiments. Major Reid, C.E.P.R.E., Farnborough, has been most helpful in securing the loan of a Bren-gun model. Mr W. H. Payne, Production Director of the Army Kinema Corporation, has patiently assisted me in the selection of the film and has also helped in obtaining War Office permission to use the film for research purposes. Finally, I offer my thanks to all student volunteers who have readily co-operated in this experiment.

Appendix

The Bren light machine gun

The Bren 303 light machine gun is a weapon which is reliable in action, simple in construction and easy to maintain. The Bren uses ·303 rimmed ammunition.

The Bren conforms to the usual sequence of actions in all automatic weapons, the actions on FORWARD movement being Cock, Load, Lock, and Fire. The actions on BACKWARD movement are Unlock, Extract, Eject, and Feed.

Let us begin with COCKING, the first of the actions on forward movement. The gun is said to be cocked when the pressure of the return spring keeps the piston extension in engagement with the sear, the nose of which protrudes up into a cut-out bent in the base of the piston extension. The sear forms part of the trigger mechanism; pressure on the trigger will rotate it, disengaging its nose from the bent in the base of the piston-extension.

LOADING is carried out when the forward movement of the breech block on the piston extension drives a cartridge from the magazine and pushes it into the chamber.

After loading, LOCKING takes place during continuous forward movement of the piston extension. The cam surface on the rear of the piston extension forces the breech block upwards. Forced upwards, the breech-block engages the locking shoulder in the body of the gun. Through this action the breech-block is sealed.

With locking completed, the cartridge can be FIRED. This happens at the end of the forward movement of the piston. The piston post, working inside the breech-block strikes the firing pin. This fires the charge.

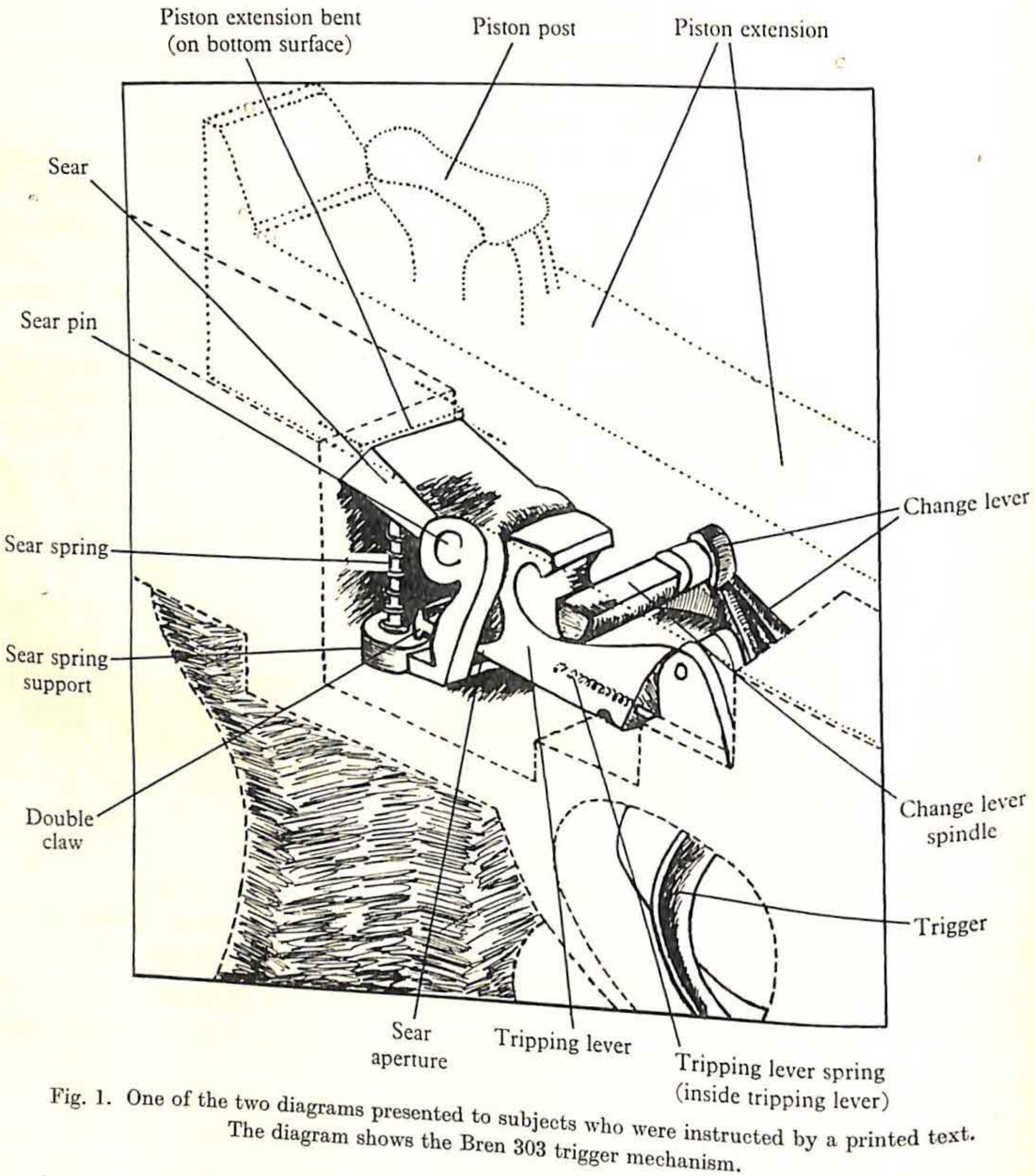


Fig. 1. One of the two diagrams presented to subjects who were instructed by a printed text. The diagram shows the Bren 303 trigger mechanism.

The forward movement cycle is thus completed and backward movement can begin. The power for backward movement is provided by the gases, generated by the explosion of the cartridge-charge. Some of the gases, which project the bullet forward and out of the bore, are diverted through the gas vent into a chamber at the base of the gun, where they meet the head of the piston. The piston is thus driven backwards.

UNLOCKING can now take place. The rear inclined face of the piston post bears against the interior of the breech block, causing the rear of block to be drawn down, thus unlocking the breech.

With the breech unlocked, EXTRACTION of the fired cartridge is effected by the action of the extractor. The extractor, a small clawlike protruberance of the breech-block, holds the case against the face of the breech-block and withdraws it from the chamber.

In the course of the withdrawal the rear of the spent cartridge case strikes the tip of the EJECTOR. This forces it away from the breech-block and the case drops out of the gun.

With the empty case ejected, FEEDING takes place. The downward pressure of the spring in the magazine first places a cartridge in position.

Meanwhile the recoiling portions continue to the rear completing the compression of the return spring until the piston extension strikes the piston buffer. This completes the backward movement and the recoiling portions move forward under spring pressure to the cocked position.

Although the Bren gun allows for automatic firing it can also be set to fire single rounds like an ordinary rifle. To see how these settings are effected, a closer examination of the trigger mechanism is necessary.

The trigger mechanism

The sear has a spring, pressing it upwards and is connected then to the trigger by the tripping lever. The tripping lever also has a spring, which keeps one end of the lever pressed upwards against the change lever. The other end of the spring connects with the trigger, keeping it pressed forward.

The spindle of the change lever determines the position of the tripping lever. The spindle has three varying surfaces: a large flat surface, a small flat surface, and the rounded surface of the spindle itself. When the spindle is rotated, it will therefore position the tripping lever at one of three different levels: large flat—lever at its highest; small flat—lever central; and round side of spindle—lever at its lowest.

The tripping lever passes through an aperture in the sear and has a double claw. With the change lever set at repetition (single rounds), the large flat surface of its spindle positions the tripping lever so that its upper claw engages the upper edge of the aperture in the sear. Pressure on the trigger will now rotate the sear and depress its nose against the sear spring. However, if now the head of the tripping lever is pressed down, its claw will disengage from the upper edge of the sear aperture. The sear spring will return the sear to its original position. Before the claw can re-engage, the trigger must be released and pressed again.

How does the trigger mechanism control the movement of the piston extension? We saw that in the cocked position the cut-out bent of the piston extension is cocked on the sear nose. When the trigger is pressed, the sear is rotated and thus disengaged from the piston extension. As the piston extension moves forward, its rear will press down the tripping lever head, thus releasing the sear from the tripping claw.

When the piston extension comes back again it first overrides the sear and then, driven forward by the return spring, returns into its original engagement with the sear nose. Every time we wish to fire a round the trigger must be pressed and released again.

Now let us set the change lever to SAFE. This brings the small flat on the change lever spindle into contact with the tripping lever, pressing it slightly downwards and positioning the claws in the centre of the apertures in the sear. When the trigger is pressed, the claws, positioned in the centre of the aperture, cannot engage the sear and it is therefore impossible to fire the gun.

Next let us see what happens when the change lever is set to AUTOMATIC. The rounded surface of the spindle now meets the tripping lever and presses it further down. When the trigger is now pressed, the lower claw engages the lower edge of the sear aperture. The low position of the tripping lever also removes the head of the tripping lever well below the level of the piston extension. Therefore movement of the piston extension does not affect the tripping lever and the gun will fire as long as the trigger is pressed and there are rounds in the magazine.

Recapitulating then: change lever set at safe, trigger inoperative. At REPETITION the sear engages each time the piston returns. At AUTOMATIC the sear remains depressed as long as pressure is exerted on the trigger.

The description of the trigger mechanism is complete. It is, however, possible to mishandle the change lever and cause a stoppage. If the change lever is set at repetition, or safe, and with trigger pressed it is then set to automatic, the arm of the tripping lever with the double claw is depressed below and in front of the aperture of the sear. On releasing the trigger the claw fouls the base of the sear and the action becomes inoperative.

To start firing again, raise the tripping lever to the middle position by setting the change lever to safe. With the trigger still released set the change lever to automatic. The claws of the tripping lever will now engage the lower edge of the sear aperture and trigger will operate correctly.

REFERENCES

BARTLETT, F. C. (1932). *Remembering*. Cambridge University Press.

JASPEN, N. (1948). *Technical Report SDC* 269-7-17. Pennsylvania State College Instructional Film Research Program.

JASPEN, N. (1950). *Technical Report SDC* 269-7-11. Pennsylvania State College Instructional Film Research Program.

LANER, S. (1954). The impact of visual aid displays showing a manipulative task. *Quart. J. Exp. Psychol.* VI, 95-106.

NELSON, H. E. (1949). The relative contribution to learning of video and audio elements in films. *Progress Report, no.* 13. Pennsylvania State College Instructional Film Research Program.

NELSON, H. E. & MOLL, K. R. (1950). Comparison of the audio and video elements of instructional films. *Technical Report SDC* 269-7-18. Pennsylvania State College Instructional Film Research Program.

VAN DER MEER, F. W. (1950). Relative contributions to factual learning of the pictorial and verbal elements of a filmstrip. *School Rev.* LVIII, 84-9.

YALE MOTION PICTURE RESEARCH PROJECT (1947). Do 'motivation' and 'participation' questions increase learning? *Educ. Screen*, XXVI, 256-9, 274-83.

ZUCKERMAN, J. V. (1949). *Technical Report SDC* 269-7-4. Pennsylvania State College Instructional Film Research Program.

(*Manuscript received* 20 *August* 1954)

FREUD AND THE CULTURE-PSYCHOLOGISTS

By TERENCE WHITE GERVAIS

I. *Introduction* (p. 293). II. *Merits and defects of both contestants* (pp. 293–295). III. *The necessary subtlety and variety of psychological method* (pp. 296–297). IV. *Boy and girl in Oedipus situation: part and whole: degrees and contexts* (pp. 297–299). V. *Psychology and the psychologist's personality* (p. 299). VI. *Freud's cultural environment: his view of woman* (pp. 300–301). VII. *Ambiguity of the terms 'social' and 'cultural'* (pp. 301–302). VIII. *Sexological aspects* (pp. 302–303). IX. *Conclusion* (pp. 304–305). X. *Summary* (p. 305). *References* (p. 305).

I. INTRODUCTION

Two distinct yet related complaints tend to be made about Freud, on which we shall here concentrate. One reproaches him for ignoring 'cultural' and 'social' factors; while the other stresses the limitations of Freud's own culture and environment.

In accordance with the method which uses every psychological concept, theory, or relation as a pattern which can *possibly* be applied to every other, we may employ four main types of analysis. In these we hope fruitfully to apply:

(1) Freud's own patterns and standards of thought to Freud himself and his work.

(2) Those of the 'Socio-Cultural Psychologists' to Freud and his work.

(3) Freud's patterns and standards to the socio-cultural psychologists and their work.

(4) Their own patterns and standards to the socio-cultural psychologists and their work.

II. MERITS AND DEFECTS OF BOTH CONTESTANTS

What we have called the socio-cultural standpoint is a portmanteau term for attitudes of many thinkers, mainly American, but including some Central Europeans. Names such as Horney, Sullivan, Fromm, couple with those of the neo-Adlerian schools; while Jung's critique of Freud touches these at minor points but suggests—if only in certain aspects—a union of the best in both viewpoints. Jung in any case stands apart through his deeper and richer interpretation of the terms 'social' and 'cultural'. Ironically the very words 'social' and 'cultural' are ambiguous, sharing an embarrassing width of meaning. Not only do the various critics of Freud differ in their use of these terms, but each single critic tends either to confuse their uses in his own thought, or else to proffer a single, lopped-off definition in sweet obliviousness of the meanings offered by his colleagues. Hence the display of this diversity, and in a sense the methodological confusion it causes, is *itself a datum* for our discussion. This confusion and diversity can, moreover, be held both in Freud's favour, as showing the unscientific methods of his critics, yet also in favour of a still more serious socio-cultural approach, one which will do justice to the complexity of the facts. The lop-sided interpretations given by several separate critics work, however, still more definitely against the socio-cultural standpoint. Social and cultural, moreover, are themselves often used as synonyms—a confusion all the more dangerous when unconscious.

Yet Freud's critics have some justification. A self-centred, obsessional way of thinking was an inevitable result of the novelty of psychoanalysis and of the insane opposition it received. Let these critics try to mould a complex set of theories dealing with vital issues, and they might well be more smug than Freud, whose creative fire kept him sensitive to many awkward aspects of truth ignored by the Yes-men. ('Moi je ne suis pas un freudiste', he said to Hanns Sachs.) But these excuses do not destroy the opposition case. Even Brierley (1951) concedes that '...very few psychoanalysts...' have paid 'attention to the second task, that of implications'. Here both sides fall into a parallel error: that of hitting on a valuable aim, while realizing it in a fashion at once confused, limited and distorted. For the culturo-social psychologists themselves fail to see the full implications of Dr Brierley's 'implications'; they interpret the term in a restricted sociological sense, a sense which even 'culture' has to bear, in accord with certain traits of their own age and environment. They have neglected to correlate psychoanalysis with philosophy and religion, and with other attitudes to truth and significance: Jung, however, has here done rich work.

Freud's own position is curious: in his later works he discusses the 'implications' of psychoanalysis, yet with a strange oscillation between admitting he is not now writing as a clinical scientist, armed with precise proof, and claiming that he is so writing. He never resolved the conflict between the anti-philosophy slogans of his 'scientificist' epoch and his own need to write what was in fact philosophy. This confusion remains as a psychological case in the annals of psychological thought. It hampered his thinking on these wider issues, and withheld him from sensitive collaboration with co-workers in philosophic and related fields (much might have emerged from co-operation with philosophically-minded historians). If his bias against philosophy was to some extent academic, that against religion was more vital. By merely *comparing* certain neurotic states with carefully chosen religious or semi-religious phenomena, he claims to shatter man's religious structure. How unscientific a boasted scientist can be! The same treatment, moreover, would with equally false logic shatter morals, philosophy, and Freud's beloved art and clearly essential science.

His cultural limitations come out in a further religious problem: his relative ignorance of all faiths save the Semitic and the Hellenic. Many details, like the 'Father'-obsession, fit still less into Hinduism, Buddhism, Taoism, etc., than into more Western faiths. This is an argument, not only against Freud's views on religion but—in a sense—against the claim that his system is universally applicable. Yet his system may be thus applicable in its clinical and purely 'psychological' aspects: are the individual and family essentially different in the East and in the West? The adjustment of such details could prove fruitful, as also the application of clinically derived concepts to the *differences* between Eastern and less Eastern religions, though without Freud's atheistic rashness.

His uncritical acceptance of the mechanico-determinist urge of his age (one of whose tenets was to explain man by his environment) leads to a stange paradox. We cannot have it both ways, and blame Freud equally for his neglect of environmental factors and for his obsession with the mechanico-deterministic views of his time. The solution to this paradox is, as often, to *recognize degrees*. We can still hold that Freud was restricted *to some extent* by the prevailing *Zeitgeist*, while admitting also that he did, in his theories, appreciate consciously the influence of environment, again *to some extent*. Whether the 'extents' in either case were just and valid is another point; we should be inclined to attribute such validity to the second issue (that of Freud's theoretical admission of environmental influence)—especially if we concede that he was concerned mainly to show firmly the basis of well-nigh universal psychological patterns. As to the first issue—his restrictions to the current *Zeitgeist*—we must concede this, too, in certain contexts: his determinism (which, again, was partly—though not wholly—a badly expressed *extension* of his necessary scientific method); his Victorian views of woman, and so forth. But Freud was not wholly of his epoch: witness the hysterical opposition to his work, his ignoring of the Hegelian identification of history and thought, or history and morals, and his lack of all naïve 'evolutionary' optimism. This belief in human irrationality and imperfection partly links him with Existentialist trends, though here his scientific balance did forbid him any extremism. As regards his tactlessly repeated formula of man being 'nothing but...' (*nichts mehr als*), it is once more a question of degree: it is futile to insist that Freud *always* said this or that, or *merely* that or this, or took one extreme view rather than another. The inhibition either against holding viewpoints 'in such or such a degree', or against admitting that another thinker has done so, is a vice of our time (from which Freud is not always free): here indeed is a context where the argument of the restrictions imposed by *milieu* and *Zeitgeist* can be applied, especially to certain of Freud's critics.

We can thus partially agree with Thompson that: '...normal self-assertion, i.e. the impulse to master, control, or come to self-fulfilling terms with the environment, was not especially emphasized by Freud', and that: 'Freud's turning away from environmental factors, as a major concern, to organic constitution was, however, most unfortunate. For this reason, he came to minimize what actually happens between people, failed to take into consideration...that it is the dynamic interaction between people which provides the locus of functional mental illness' (Thompson & Mullahy, 1952, pp. 13–27). Once more, we would agree only up to a certain point: for Freud hardly neglected inter-human relations and influences—the whole theory of child-psychology, of trauma, repressing of the Oedipus complex, or the super-ego, disproves such an accusation. A trauma, a repression, a complex, is caused by an event or by events in which usually some other person or persons take a vital part. While Freud's conception of the super-ego is not always as flexible as the facts demand, it remains a recognition of a 'social' and 'cultural' influence. His chief concern, however, was admittedly with individuals and with their effect on each other: his conception of the 'social' lacked the herd-element stressed by many of his critics. Not so easily as many of them did he fall a victim to the identification of more or less intimate personal relationships with larger groups and contacts which alone deserve the full epithet 'social'. Even in dealing with close personal relations, however, he tended towards two limitations: he concentrated perhaps overmuch on relations derivable from or parallel to those clinical contexts from which psychoanalysis takes its origin, and he seemed less aware of the infinite variety of *persons* than of the nature of their influences as reduced to generalized *patterns*.

This preference for 'patterns' over 'persons' we must not, however, exaggerate: first, because Freud himself was often aware of the capricious variety of personal individuality, and because a certain degree of 'depersonalization' is inherent in all science. Certainly psychology is in a different position from other sciences as regards this necessity of method, for its theme is the human person; but the scientist is also a human person—hence a labyrinthine problem of *possible* subjectivism, and of the twistings of the 'Selfhood' of both thinker and theme, to which no other science offers a parallel in any comparable degree.

III. THE NECESSARY SUBTLETY AND VARIETY OF PSYCHOLOGICAL METHOD

We have here reached a point which is inherent in many of our previous arguments. Is there not a salient difference between explicitly ignoring the importance of socio-cultural factors (or even denouncing their importance), and merely stressing the personal, familial, and sexual factors which must in any case prove the *basis* for all fruitful discussion of the socio-cultural field itself? And this, in the main, is what Freud has done, though not always adequately. Historically his approach was, on the whole, the only suitable one for the nascent science of psychoanalysis. What is important is not to regard the two approaches as opposites: too much thinking is done (and sometimes by Freud himself) in naïve oppositional terms. Not every situation is best analysed into a *duality* (though here again Freud was a frequent offender: witness his too numerous antitheses of instincts, etc.).

Of course everyone will draw the line differently between methodological necessity and its extension into an overwhelming doctrine; the method cannot help being three things: a method as such; a part of the very subject-matter; and, to some extent, a potential doctrine. As regards this potentiality of doctrines must we not remember that different thinkers will each tend to regard a method as having become a doctrine or not, in different contexts, according largely to the psychology of these thinkers themselves? If Freud, for instance, is deemed too 'subjective', each of his critics will still act according to his own subjectivity in attributing such a quality to Freud in a wide variety of contexts and degrees. So the 'psychology of the psychologist' hunt, while a most essential aspect of a full scientific method, provides its own problems, and apparently insoluble circles. The whole problem of the inter-relation of a method both with subject-matter and with doctrine is, moreover, one facet of the vast pattern of 'Selfhood' (*Selbstheit*), of Presuppositions, Identities and Tautologies. This asks: what is the relation between anything, fact or idea, and any of the following: its existence, its expression, its emphasizing, its parts or aspects, its relationships, its 'defence', and our consciousness or not of it? Freud cannot be blamed for such a ubiquitous pattern of life's dilemma. And this applies, not only to the relative ignoring of the social factors in favour of basic individual patterns, but equally in the cases of 'depersonalization' and of 'determinism'.

Moreover, the determinism, the depersonalization, or the individualist stress may or may not overstep its valid sphere, but seldom entirely so: hence the need to think in terms of degrees, and also of particular contexts. Here, by the way, is a sphere where the methods of Statistics may be of great use to psychology; or rather, perhaps, statistical methods and statistical 'mentality'. Instead of claiming the supreme (if not sole) validity in psychological method, let these act as shrewd discerners precisely of the *degrees* of fact or validity, and of the exact *context*, to which such-and-such a theory or critique may apply. An unusual role, this, for statistical mentality and method, but surely a useful one: at once a subtilizer and a safeguard.

Another valuable principle is what may be called the pattern of the 'Two Ends'—by which each term in a two-termed situation may be considered, according to the context, as the origin, means, or aim, or as the cause or effect of the other, and the general situation described equally well with either term as the main factor. This pattern (which is of course not universally applicable) can often save us the futile effort of warring one-sided dogmatisms: in art, for instance, with factors such as unity and variety; restraint and intensity; a 'straight line' aiming at a final 'curve', or vice versa; passion and serene form; individual and 'typical' forms; subjective and objective elements, etc. Similarly, in psychology. Can we not agree with Fr. Victor White, that: '. . . even Professor Flugel must assume (for he cannot possibly prove) that "God" is a projection of the super-ego rather than allow that the super-ego might be an unconscious introjection of "God" (or at least of certain divine images and attributes)'? (White, 1952, p. 20). In other words, we have here a pattern of the 'Two Ends', which may well fit also, *mutatis mutandis*, into the problem as to the primary of 'individual' or 'socio-cultural' factors (or even as between 'socio-cultural' and 'sexual' factors). If Freud has not seen this, neither have most of his opponents.

Another too sharp antithesis is that of 'subjective' versus 'objective' methods in psychology. We have already shown how, in psychological as against other forms of science, the subjective forms a large part of the objective theme and data themselves, and how an awareness of such subjective data is essential to the psychologist, who—to complicate matters further—adds to the labyrinth of his own subjective slant. Hence, though complete 'objectivity' is the scientific ideal, in psychology at least this will involve a strong subjective element, both an awareness of subjective data and a subtle use of one's own subjective attitude in discussing such data. The 'discovery of general laws of Mind' (objective) and the 'description and understanding of the unique undivided personality' (subjective) are both needed; their results can '. . . be correlated, and can be used to correct each other' (Pear, 1948). From another angle, 'What is experienced as meaning can be objectively described as a functional integration of mental processes' (Brierley, 1951, p. 94). Why, now, trouble to agree whether the one type of psychological method is either (*a*) 'prior', or (*b*) 'superior' to the other? In such a case, the acceptance of the 'Two Ends' both clarifies and enlarges our approach. On the whole, we find a greater awareness of such breadth of method in Jung than in Freud.

IV. BOY AND GIRL IN OEDIPUS SITUATION: PART AND WHOLE: DEGREES AND CONTEXTS

Freud's work, however, is full of patterns of thought and fact, some of which can even—by a sleight of method well adapted to psychological science—serve to indicate solutions of wider psychological problems. Can we, for instance, not use his distinction (a subtle piece of analytical insight) between the *boy's* and the *girl's* early routes through the Oedipus complex and other familial relationships, as a pattern which can help analogically to 'solve' the dilemma of 'individual' versus 'socio-cultural' approach, or that of 'subjective' versus 'objective' methods?

Freud shows how the pattern of child-parent relationship differs in the case of a boy and of a girl—the girl's relationship having an *extra* stage: the Oedipean attraction of the boy to his mother, and of the girl to her father is preceded in the latter case, by the girl's attraction to her mother. Hence the girl has two stages to the boy's one. Is it too fanciful to suggest the analogy that we shall regard the 'individual' and the 'socio-cultural' approaches, in *some* contexts as methodologically 'equal' (like the earliest relations of both girl and boy to the mother, or their later respective attachments to parents of the opposite sex) and in *other* contexts—probably the more important kind—as conforming to the 'unequal' pattern just mentioned? In other words, let us realize that 'socio-cultural' psychology *involves* a 'previous' stage of individual psychology, and that this stage, moreover, is not packed into the former and wholly absorbed, but continues ceaselessly to act and react on the very texture of the socio-cultural factors and analyses themselves. Similarly, we may sometimes treat the 'objective' and 'subjective' methods as methodologically 'equal', but also realize the 'twofold stages' in the 'objective' approach, which contains a 'previous' subjective phase which continues its influence. To a slightly lesser degree we may say the same the other way round, that subjective method contains an objective phase; in which case, we have yet another example of the 'Two Ends' pattern, and also of the need to perceive and admit *degrees*.

At last we may seem to have reached a logical and fruitful *modus vivendi* between the 'individual' and the socio-cultural approaches. Yet the accusation cannot be ignored that Freud—or any other psychologist—may have called the part the whole and thus limited his own approach and conclusions: this accusation is indeed made *from and against all sides*.

Thus, Freud on Adler: 'From a highly composite unit one part of the operative factors is singled out and proclaimed as the truth' (Freud, 1918; English trans. 1933, pp. 526–7).

Yet Thompson says of Adler: 'Adler was the first one to point out that sexual disturbance, instead of causing neurosis, was but one of the situations where neurosis manifested itself' (Thompson & Mallahy, 1952, p. 159).

We indeed seem in the presence of some hyper-flexible inside-out indiarubber of the psyche, where, to some degree and in some contexts, *A* may be a part of the whole *B*, while, in other contexts and again to some degree, *B* may be a part of the whole *A*. In psychology, as yet so unformed a science, this may not be so absurd as it seems. The pattern of the 'Two Ends' may legitimately be applied here, though with certain caution. Notice our reservations, 'to some degree and in some contexts'; we may further apply these reservations to the validity of the 'Two-Ends'-pattern itself as regards this problem. For this pattern is by no means the sole one in which the problem can be stated, let alone clarified: we may, for instance, well consider it in the light of an extended pattern of interacting Wholes, *Gestalten*, and of different levels or degrees of micro-macroscopic scale. Moreover, the ubiquitous design of 'Selfhood' and Pre-suppositions, to which we have already referred, appears again in this problem; for so much depends on the pre-suppositions made by both Adler and Freud in their major terms and theories: 'Sex' in the more ordinary sense is certainly *a part* of the wider causes, and aspects, stressed by Adler and Jung; yet, in Freud's own sense, it *includes* many of these, and in this way Adler's 'Power'-theory may seem a distorting extraction of a single element. Yet Adler in turn, in opposing his master's 'narrowness', pre-supposed a number of important factors in the world, which he assumed he had sufficiently brought to light in outlining his alternative concepts. In both cases, certain aspects of life are assumed, and either not mentioned explicitly or else *included* in some wide concept, an arrangement, however, which does not fit in with some other thinker's assumptions, inclusions, conscious and unconscious use of terms and so forth. In these cases we really do seem to have a design of interlocking 'Wholes'; the interlocking, however, is not in the least simple; it is a twisting in-and-out with continually changing emphasis and validities.

Yet none of these solutions may entirely satisfy us. Both in this Freud–Adler contest and in that between Freud and the socio-cultural standpoint, we surely need a method which will sift out clearly *to what degrees and in what contexts*:

(1) The 'Two-Ends' pattern is valid. (This would seem to be the case, in the main, in the discussion whether anxiety causes neurosis, or neurosis anxiety—a point on which Freud himself fluctuated.)

(2) The patterns of interlocking or otherwise related wholes and *Gestalten* and of different levels or degrees of micro-macroscopic scale, are valid, or at least useful and suggestive.

(3) The pattern found in the boy-and-girl Oedipus histories is valid, or at least suggestive and useful.

(4) None of the above really meet the problems. One attitude or the other (i.e. Freud's or Adler's, in the one case; that of Freud or that of the socio-culturalists, in the second) must have the victory.

Research into such questions of degrees and of contexts forms surely a valuable, and fairly novel, type of psychological method.

V. PSYCHOLOGY AND THE PSYCHOLOGIST'S PERSONALITY

In contrasting Adler and Freud we have touched on yet a further problem: the conflict of personalities within psychology itself. This is a result of that *mélange* of scientist and subject-matter in which psychology is unique. Every psychologist tends from time to time to criticize his opponent's views on the grounds of the latter's own personal psychology; yet he will deny doing this, and strenuously object when the retort discourteous is made to him. Then it becomes 'unscientific'. Again it is difficult but imperative to find the *juste milieu* between ignoring and over-stressing the (probable) subjective motives for any psychologist's opinion. Many objections to Freud's theories, based on what was supposed to be his personal character, fail in this respect; so also do many of Freud's own replies to criticism, though here the inane and insane nature of the world's first reaction to Freud provides an obvious extenuating cause. Such an excuse, however, scarcely applies to his attitude toward Jung and Adler; here he seems to indulge in an attack half savage, half tired-ironic, on the inner motives of the 'renegades'. Yet when Adler suggested that the founders of psychoanalysis be analysed, and analysis applied to the creation and holding of analytical theories, this was regarded as lèse-majesté! Moreover, does not Freud, in his *Autobiographical Study* (1935), confess to a power-complex, to jealousy, and to the periodic need to quarrel with a close associate? Yet such factors do not affect Freud's own positive theories and insights.

Ernest Jones, defending the work of Melanie Klein but analysing the opposition it aroused, describes the facile instinct to place her novel theories in the same category as those of Adler, Jung, Stekel and Rank, who were 'influenced by subjective motives—a rationalization of inner resistances.... Yet, if psychoanalysis is to remain a branch of science it is evident that...advance beyond the limits he (Freud) reached is inevitable...' (Jones, 1948). As if Jung's mentality can ever be a more subservient derivative or decoration of Freud's! The ignorance displayed here of Jung's *type* of work, so often complementary rather than opposed to Freud's, is embarrassing but typical: is there no 'rationalization', in Dr Jones's own case, of some 'inner resistance'? May we not extend his last sentence, to the effect that 'if *psychology* (not merely psychoanalysis) is to remain a branch of science,...advance beyond the limits he (Freud) reached is inevitable'?

Psychology is wider than psychoanalysis.

VI. FREUD'S CULTURAL ENVIRONMENT: HIS VIEW OF WOMAN

Let us turn to Freud's own cultural environment and attainments: we feel that these are less wide than Jung's, but far richer and broader than those of most anti-Freudians.

In his *Autobiographical Study*, Freud rebuts the Sorbonne professor's gibe that 'psychoanalysis is alien to the *génie latin*', by asking if it were the favourite child of the *génie teutonique*; and his irony 'defends' Vienna from the charge of contamination with his work. Yet the Viennese *milieu* may have stimulated him, with the 'psychological' and cynically introspective slant of its literature, with its melting-pot of nations and types, with its blend of neurosis and gay charm, its relatively free expression both *of* and *about* sex, and the *Schlamperei* which bred a half-tolerance lacking at that date either in Protestant lands or in the full, sharp Catholic South. Precisely the 'universal' traits in Vienna helped Freud indirectly; 'localist' criticism is thus in error.

Behind these reproaches lay an often subhuman antisemitism. Freud, for all his wide disinterestedness, curiously agreed with his insulters in declaring: 'It is, perhaps, no coincidence that the inventor of psychoanalysis was a Jew', stressing the Jew's acquired power of standing alone. He might have added an absorption in, yet detachment from, prevailing cultural forms, the ability to sift the best in each of them, and to embrace the most diverse facets of the human spirit. Perhaps less admiringly we might point to the overwhelming role of the Jewish family- and Jehovah-traditions in Freud's extreme stressing of the *Father*-complex; while his blind negation of religion may be due in part to this rigid faith, and his rejection of it in the fire of his scientific knowledge. Towards Christianity he felt an ambivalence of intellectual interest and emotional repulsion, while Oriental faiths seem closed to him. Thus it is hard to explain Fromm's statement that Freud's work was the typical result of a 'Protestant environment'; that religious environment was Jewish and Catholic. Why thus bolster a case at the cost of obvious distortion?

No less and no more valid than critiques against Freud's Jewish background and nineteenth-century scientific rationalism, were his own references (especially after their quarrel!) to Jung's Protestant-pastor background and to Swiss mentality in general; or his attacks on America, through partly well-built fears lest it cheapen his psychoanalytical creation; while in and after his lectures at Worcester, Mass., his first public exposition outside Central Europe, and first official welcome in any University, he saw little in America to admire. A grudge-bearing lack of gratitude, indeed not without causes from his hard early life, is hinted at in the *Autobiographical Study* itself.

Of greater strength are the critiques which expose Freud's limited view of woman; even according to Brierley, it is the weakest part of his work. Witness his Victorian views, his uncritical acceptance of the male 'order' of our culture, his equation of the ideas that woman is *different* and that she is *inferior*, his frequent confusions between the 'masculine' and (what is thereby meant) the 'human' ideal; his references to criticisms made against women without any account of criticisms levelled at men; his assertion that child-bearing, house-slaving 'enigmatically realistic' woman is 'less ready to submit to the great necessities of life' (=men?); or the blank equation of woman's subjectivity with a lower ethical level, omitting reference to the kindred feminine virtues. Nor can he conceive that certain female traits are due to age-long social conditions imposed by males. Here we can definitely accuse him of underrating or ignoring the 'social' aspect of psychological fact. He talks of 'feminine vanity', ignoring the male kind. To a female psychologist (and to most women) this lacuna itself shows male vanity. Yet he refers to the mother-son relationship as 'supreme', and we know from his own words how strong was his bond with his mother. (All this is not to deny his many discoveries in female psychology, e.g. penis-envy—though women psychologists like Deutsch and Horney (and even males like Jones) have often been able to adjust certain details.)

These arguments as to his 'masculinity-complex' he would combat by saying that 'the use of analysis as a weapon of controversy obviously leads to no decision' (Freud, 1931, English trans. 1950, p. 258): in other words, the masterful male can apply the 'personal' argument to his female and feminist critics, who cannot return the compliment. His own use of such a method (as applied even to males like Jung, Adler, Rank, Stekel), and his dismissal of it as 'unscientific' when used against himself, might even show—on his own terms—a 'feminine' trait, rather than the impartial male Olympian search for truth.

Truly it is difficult to draw a precise line between ignoring the value of this 'subjective' method; in dealing with woman, however, Freud does not even attempt the task. Similarly, it is hard to draw an exact line between a *recognition* of the masculine bias of our society, especially when this touches female development, and a *psychological theory* based on the assumption that such a bias is inherent in the order of things. On the whole, Adler and Jung both emerge from this test with finer sensibilities than Freud.

This is perhaps the sole exception—judged by his own standard of piercing insight—to the liberal air which pervades Freud's work. His maligned epoch, with its urbane civilizedness and growing social conscience, tinges this work with its 'Live and let live': is this inferior to our strange blend of increased specialization and mania for the 'totalizing' of all activities—even of those ideals of intellectual truth, which in Freud's day stood unquestioned? Harsh pragmatism, either of political ideology or of (partly valid) 'total social criteria', is no advance on at least the attempt at scientific integrity. Hence, by the culture-psychologist's own criterion, we can judge Freud—though but partially—through his cultural air and aim: and these were in many ways more humane and more scientific than our own.

VII. AMBIGUITY OF THE TERMS 'SOCIAL' AND 'CULTURAL'

Perilous ambiguities, moreover, beset the very terms 'social' and 'cultural'. Each is used in a variety of meanings, to suit the context, and furthermore 'cultural' and 'social' are confused with each other. Such laxities cause error and misunderstanding.

Psychologists and non-psychologists alike confuse the term 'culture': witness the debate whether culture should be primarily 'social'—a nice picture of co-operative redoubled muddle. 'Social' and 'political' are also identified, because in *some* contexts they may bear similar or parallel roles. Adler himself uses 'social' now to mean political, now merely gregarious, and now as regards intimate relationships. His system, despite its many insights, overstresses the 'social' aspect, due partly to Adler's own belief in socialism, and even to the word-play contained in the term: here we must to some extent use the social criterion in judging the psychological theory of social adjustment itself. Again, 'The inferiority complex...finds its explanation in the relatively greater deficiency of social feeling' (Adler, 1936, p. 115): here 'social' bears yet another nuance; but such nuances are not recognized, ordered, or allowed for. (Cf. also Brachfeld (1951): '...as Adler said, sex is not just a private matter'.) As to 'culture', must we assume that Freud ignored the Greek art and thought enshrined in Christianity, when he described the latter as 'a cultural regression as compared with the older Jewish religion'? (Freud, 1939, p. 142). In other contexts (cf. our later discussion of Incest) the anti-Freudians fall into still louder muddles of 'cultural' terminology.

Timely is Lindner's attack (1952) on the 'adjustment'-mania which destroys adventure and protest: yet here the culturo-social aspect strangely intervenes, for 'social adjustment' is more deified in Dr Lindner's country than in Europe. Moreover, the 'social' tinge common in American 'individualism', the stress laid on a person's 'fulfilling his self' in the interests of a healthy society of similarly fulfilling persons, accounts for the unconscious intrusion of 'social' proposals like Dr Lindner's crusade for eugenics.

Another slant on the socio-cultural aspect of psychology appears in Jung's sense of *the continued coexistence of various cultural attitudes of the past* ('There are people who, psychologically, might just as well have lived in the year 5000 B.C....', while 'a few of us belong to the 3rd or 4th millennium A.D....') (Jung, 1952). This contradicts the facile assumption that every 'main region' of the world has, at a given time, *one* 'culture'. This defect both of vision and of method is parallel to that which speaks incessantly of 'society': as if there were not (even for each individual) myriad interlocking, and often conflicting, communities or groups that link each of us with other humans (even a relationship *à deux* or *à trois*, or even that between a dead scientist or artist and all who benefit from his work, can still prove a potent 'society').[1] How perilously vague, then, to assert that 'art must reflect society...social values must be paramount....Psychology should stress the social angle, and is correctly a branch of social studies'. *We must, in fact, 'socialize' the idea of 'society' itself: this is a methodological parallel to applying Freudian analysis to Freud's own theories.*

VIII. SEXOLOGICAL ASPECTS

Relevant to the socio-cultural argument is Freud's sexology. His emphasis on sex has many facets and causes: scientific precision; reaction against the anti-sex conspiracy; the difficulty of expression in a new science; and the general *Zeitgeist* of an era at once romantic and determinist-materialist. In that era we still largely exist; it is a half-truth to explain away Freud by the sexualism of his times, for that sexualism has not abated, and we should beware of rationalizing an inner resistance to our own natures. It is inaccurate either to play down the sex-motif (like many modern psychologists), or to regard sex-interest as 'the typical neurosis of modern Western man' (Jung, 1928, p. 34): as if other ages of the West, or any Eastern epochs, have not shown similar traits! Hinduism offers but one of many potent examples. The low role of women in most Oriental cultures to some extent confirms, as well as complicating, this point; while their seclusion has as both cause and effect an intense sexual valuation, curiosity, and jealousy; and these are found to some degree in the relative segregation of sexes in Southern Europe, where surely sexuality is as active as in the North. All this implies that we should examine afresh every *detail* of sexualism, which may well differ in various *milieux*: yet not enormously enough to upset the basic teaching of Freud.

Again, the full or partial seclusion of women is balanced in many lands by a more general homosexuality. To ignore this facet shows an unscientific prudery, a semi-neurosis from which Freud was free. To him homosexuality was as much 'sex' as the more normal form, not did he judge its importance or values by such criteria as social adjustment, benefit to 'the' community, wholesome or unhealthy moral tone, or military effects. His frank analyses of homosexual elements aroused unconsciously still more opposition than his other sexual doctrines. In this he is more objectively scientific, and less conditioned by culture and social environment, than his opponents who blame him for being too preoccupied with sex. (He himself might have analysed this very resistance in terms of an unconscious homosexuality.) Nor—to mention one facet of a vast problem—can we ignore the (anti-) homosexual complex so evident in the United States and in Britain, to explain the attitude of many socio-cultural psychologists to homosexuality.

---

[1] For more extended treatment of this problem see my article 'Render unto Caesar' (Gervais, 1946) and poem *Patrick Freed* (Gervais, 1948).

Let us therefore examine *in what details* the various Eastern civilizations differ on this point from each other and from the West. (Consider, for example, the relation of the Oedipal theories to China's extreme family tradition.)

Sometimes, however, an important part of Freud's system is deemed inapplicable to either East or West. A salient case is the Incest problem—where the ambiguity of the term 'culture' only adds to the confusion, alike of data and of deductions. Do the widespread laws against incest indicate a 'natural' abhorrence of it, or rather an almost panic prohibition of a 'natural' desire? We shall cite some passages to show *two* types of confusion: (i) that in terms like 'culture' and 'intelligence'; and (ii) that involved in any attempt to generalize human history in such complex problems.

Dalbiez (1941) states, for instance, following Malinowski: 'The family is the cradle of culture.... Hence...incest is seen to be unreconcilable with the very existence of culture.' He then quotes Malinowski: 'Culture in the form of law, morals and custom—forces the male...to keep guard over the pregnant woman.' Is this the 'culture' implied by learning and the arts? Is the Eroica Symphony or the Relativity theory irreconcilable with incest? Even apart from these questions, we dispute the naïve identification of 'culture' with certain types of culture—whereas in others incest has officially flourished (e.g. often as a safeguard for the family unit!).

Dalbiez, again, in equating 'morality' with sexual morality, is more sex-ridden than Freud, who merely showed the sexual origins, and the continued quasi-sexual aspects, of phenomena not usually classed as such. Dalbiez, while accusing Freud of not going deeply into these questions, himself ignores such dilemmas as the collisions, in certain contexts, between sexual custom and other moral laws, like charity and imaginative understanding (the second-grade treatment of 'illegitimate' children is 'rooted in man's human (as distinct from his animal) nature'). 'Cultural sexual morality'—as if persons of finer aesthetic and intellectual culture were not often freer in their sexual opinions!—becomes the unanalysed fetish. Yet again, if all cultures have imposed restrictions on sex, that does not mean we can or should not urge *certain* adjustments or the removal of *some* restrictions. Moreover, each culture tends to a different sexual-moral programme: there simply has not been any such thing as one sole official attitude towards the following aspects of sex:

Monogamy, polygamy, polyandry, homosexuality in males or females, divorce, birth-control, abortion, bastardy, marriage between cousins or to a deceased wife's sister, or masturbation.—Parochialism and lack of elementary logic can go no further.

IX. CONCLUSION

'Social' and 'cultural' are thus perilously ambiguous. Yet, judging the culturo-social psychologists by their own standards, we may suggest that, if Freud is to be blamed for adhering to the materialist determinism of his epoch, so should these particular critics bear censure for being unduly influenced by *their* time and place, by the sociological craze, which—like the materialism induced by the first flush of scientific conquest—represents the exaggeration (and thus distortion) of a necessary trend. We admit the worth of the 'social' and sociological trends; yet their excessive encroachment is a defect of our time, to be corrected and compensated for with finesse, by present as undoubtedly by future thinkers.

Notable, as we have seen, is the meaning—at once mechanical and vague—given to key-words like 'society' and 'culture'. Notable, too, is the ambiguity between the meanings acted on by different thinkers, and produced unconsciously by diverse environments: for instance, the tendency (not confined to America) to identify the individual's free self-fulfilment with 'constructive service to a progressive community', colours many attacks on Freud: as if such could possibly be the essence of self-fulfilment and freedom, though it is a fruitful corollary, opportunity, and side-product. Yet these same critics (encouraged by some looseness of Freud's language) will blame him for regarding the tendencies repressed by the super-ego as in themselves necessarily bad; while we wonder what use will be found for these often desirable traits by the society of Fromm, Sullivan and others, with its background of vague idealism plus ruthless efficiency: is 'functional social utility' to be their most admired aspect? Freud at least was too 'cultured' to regard the supreme test of art, science, or morals as lying in their social value as defined by social-minded social psychologists. Neither did he restrict 'society' to one specific form, while he gave frank warnings on the nature and conduct of groups as such. A final suggestion: can we find any relevance in Zilboorg's critique (1938) of the overstressing by American psychologists of *anal* reactions, which Brierley reinforces by noting British *oral* preoccupations (Continental Europe is of course needed to complete the picture)?

Yet there are many ideas of these critics of Freud which we can, in general, accept—though, like Sullivan's theory of 'Interpersonal Relations' (1947), they can often be adjusted to the basic Freudian structure. Once more the elusive patterns of what is basic or presupposed, and the relations between method and doctrine! Moreover, in the widespread habit of posing *additions* to the Freudian basis as a *contradiction* of it (a habit encouraged by the most fanatical Freudians), is there no psychological current, conscious or not?

Part of the socio-cultural thesis may be due to an unconscious frustration and embarrassment at the dogmatic 'one-method' attitude exhibited by each rival school, and at the mania for describing man as 'Nothing but...' (*nichts mehr als*): a dogma deplored by Jung but often evinced both by Freud (often through verbal carelessness) and by the culturo-social schools. Against these tendencies, there may arise a semi-articulate urge towards the needed breadth of psychological method—in which psychoanalysis, Jung's system, *Gestalt*, and socio-cultural approaches will go hand in hand with each other and with experimental and statistical methods. If Einstein (1949, p. 684) is right in depicting the physicist as, in different facets, a realist, idealist, positivist and Pythagorean, how much more valid is such a synthesis for the complex, immature, and objective-subjective science of psychology. In this way, we curiously *sublimate* the culturo-social thesis by bringing the variety of sociality and the breadth and nuance of culture into psychological method itself.

X. SUMMARY

Freud is partially guilty of the twin sins of omission charged against him: his own cultural limitations, and his ignoring of culturo-social factors, while in some respects he shows these factors in a subtler and deeper light. Moreover, these charges have been exaggerated to a point which has led us to expose *comparable* faults in his opponents, who, on the whole, betray grosser defects in these matters than Freud himself: and all this, quite apart from the countless achievements to his credit in other branches, aspects and contexts of psychology.

REFERENCES

ADLER, A. (1936). *Social Interest.* London: Faber.

BRACHFELD, O. (1951). *Inferiority feelings.* London: Allen and Unwin.

BRIERLEY, M. (1951). *Trends in psychoanalysis.* London: Hogarth Press.

DALBIEZ, R. (1941). *Psychoanalytical method and the doctrine of Freud* (trans. T. F. Lindsay). London: Longmans.

EINSTEIN, A. (1949). In P. A. SCHILPP (ed.): *Albert Einstein: Philosopher-scientist.* Evanston, Illinois: Library of Living Philosophers.

FREUD, S. (1918, trans. by Strachey, J., 1933). *Collected Papers*, vol. III. London: Hogarth Press.

FREUD, S. (1931, trans. by Riviere, J., 1950). *Collected Papers*, vol. V. London: Hogarth Press.

FREUD, S. (1935). *An Autobiographical Study.* London: Hogarth Press.

FREUD, S. (1939). *Moses and Monotheism.* London: Hogarth Press.

GERVAIS, T. W. (1946). Render unto Caesar. *Hibbert. J.* XLIV, 263–70.

GERVAIS, T. W. (1948). *Patrick Freed.* London: Fortune Press.

JONES, E. (1948). Preface to KLEIN, M.: *Contributions to Psychoanalysis 1921–45.* London: Hogarth Press.

JUNG, C. G. (1928). *Contributions to Analytical Psychology.* London: Kegan Paul.

JUNG, C. G. (1952). Preface to WHITE, 1952 (see below).

LINDNER, E. (1952). *Prescription for Rebellion.* London: Gollancz.

PEAR, T. H. (1948). Perspectives in modern psychology. *Brit. J. Psychol.* XXXVIII. 158–64.

SULLIVAN, H. S. (1947). *Conceptions of Modern Psychiatry.* Washington: William Alanson White Psychiatric Foundation.

THOMPSON, C. & MULLAHY, P. (1952). *Psychoanalysis: Evolution and Development.* London: Allen and Unwin.

WHITE, FR. V. (1952). *God and the Unconscious.* London: Harvill Press.

ZILBOORG, G. (1938). Some observations on the transformation of instincts. *Psychoanal. Quart.* VII, 1–24.

(*Manuscript received* 18 *July* 1954)

A STUDY OF RACIAL ATTITUDES DURING AND AFTER THE WAR BY THE THURSTONE TECHNIQUE

By A. B. FITT

(*Professor Emeritus*) *Auckland University College, Auckland, New Zealand*

I. *Introduction* (p. 306). II. *Analysis of the attitude scores toward Germans, Japanese and Maoris* (pp. 306–308). III. *Comparison with a study by Remmers & Wood* (p. 308). IV. *Summary* (pp. 308–309). *References* (p. 309).

I. INTRODUCTION

Before the last war the writer had been doing some work, much of it with his class in experimental education at Auckland University College, exploring the usefulness of Thurstone's method of measuring sentiments or attitudes. For some years three attitude scales (towards Germans, Japanese and Maoris) had been used with the class for demonstration purposes. The scales had been constructed on the lines laid down by Thurstone & Chave (1929), most favourable attitude being placed at the zero end of the scale.

When the war started it appeared that there was a good opportunity of checking the validity of the technique by noting whether the attitude towards racial groups with whom we were at war changed in the expected direction, while the Maoris appeared to offer a good 'control' as against the Germans and Japanese. Unfortunately, the results of testing before the war had not been kept, as the work was of a general demonstration type, not conducted with any particular aim. Consequently, the results available for this study are those of 1940 and succeeding years. Testing was done each year, usually in the month of August, near the end of the winter term. The class used was that in experimental education with similar membership in the different years.

II. ANALYSIS OF ATTITUDE SCORES TOWARD GERMANS, JAPANESE AND MAORIS AND THEIR VARIATIONS

Table 1 presents the facts of the analysis, giving the fluctuations in the mean attitude scores. The number of cases in each group varied from 20 to 71. The mean scores of the different groups can be roughly interpreted with reference to an approximate indifference score of 5·1 which corresponds to the statement '...neither better nor worse than any other people.'

The most obvious fact of the fluctuations in attitude toward Germans is a rise toward a less favourable attitude in 1945, 1946 and 1947, that is, at the end of the war and for two years after it. The score has fallen considerably in 1950, and by 1953 it has reached the indifference zone. By this time the attitude toward Germans has returned to a normal pre-war level. The main differences in attitude, those of 1945, 1946 and 1947 when these years are compared with the low level of 1942, are the only ones with any degree of significance. The figures of Table 2 show clear significance for 1946 and 1947 and only slight significance for 1945. The unfavourable nature of these attitudes is presumably related to the disturbing and horrifying effects of the revelations of atrocities, concentrations (e.g. Belsen) and other circumstances observed after the cessation of fighting.

Study of the attitude scores with reference to Japanese shows a somewhat similar trend to the German scores. After 1941 all the values lie in the unfavourable zone (above 5·1), those for 1943, 1945, 1947 and 1950 remaining at a fairly high level. Even the 1953 value is above the indifference zone, the return to normal peace relations not being as clearly indicated as in the German scores. Table 2 shows that, of these four high scores, those for 1947 and 1950 have significant differences when compared with that for 1941, that for 1945 is almost significantly different from that for 1941, and the 1943 score lacks something of significance, although the ratio is not negligible.

Table 1. *Means of attitude scores and standard deviations*

	1940	1941	1942	1943	1944	1945	1946	1947	1950	1953
n	25	20	56	20	22	31	62	71	29	33
						Germans				
A.M.	5·4	5·3	5·1	5·3	5·2	5·7	5·7	5·6	5·3	5·1
S.D.	0·78	1·00	0·74	0·79	0·81	1·41	1·11	1·05	1·21	0·78
P.E.M.	0·1054	0·1509	0·0667	0·1192	0·1165	0·1708	0·0950	0·0840	0·1515	0·0900
						Japanese				
A.M.	5·1	4·9	5·2	5·5	5·3	5·5	5·3	5·5	5·5	5·3
S.D.	0·75	0·35	0·72	1·05	0·56	1·10	0·76	0·71	0·90	0·80
P.E.M.	0·1006	0·0524	0·0649	0·1584	0·0805	0·1333	0·0651	0·0568	0·1127	0·0900
						Maoris				
A.M.	4·8	4·1	4·8	5·3	5·2	5·5	5·0	5·3	5·0	4·8
S.D.	0·81	0·71	0·88	1·20	1·17	0·86	1·04	1·13	1·08	1·01
P.E.M.	0·1090	0·1065	0·0793	0·1810	0·1678	0·1059	0·0891	0·0900	0·1353	0·1185

We have noted a more decisive trend for the German attitude scores as against the Japanese ones to return to peace levels after the war. It must be remembered that Japan is a fairly near Pacific neighbour for New Zealanders and that a war situation comes more easily to mind with them than with the Germans. Near neighbours know a good deal about each other.

Table 2. *Significance of differences in attitude scores*

	Germans			Japanese			
	1942 and 1945	1942 and 1946	1942 and 1947	1941 and 1943	1941 and 1945	1941 and 1947	1941 and 1950
Diff.	0·6	0·6	0·5	0·6	0·6	0·6	0·6
P.E.D.	0·1833	0·1160	0·1072	0·1668	0·1432	0·0772	0·1242

	Maoris								
	1941 and 1940	1941 and 1942	1941 and 1943	1941 and 1944	1941 and 1945	1941 and 1946	1941 and 1947	1941 and 1950	1941 and 1953
Diff.	0·7	0·7	1·2	1·1	1·4	0·9	1·2	0·9	0·7
P.E.D.	0·1523	0·1328	0·2100	0·1987	0·1501	0·1388	0·1394	0·1721	0·1593

Whereas the changes in attitude toward Germans and Japanese were very much as expected, those toward the Maoris, New Zealand's early inhabitants, were little short of a complete surprise to the writer. There is certainly not the expected picture of a 'control' group. The Maori measures show the same kind of fluctuation as the Japanese and German, but at a lower (more favourable) level. The attitude scores of four years (1943, 1944, 1945 and 1947) are unfavourable, the others being favourable.

Two factors appear to have been at work here influencing the attitude toward Maoris. One, which has been noted frequently by workers in the social sciences, is a tendency in attitude toward one alien racial group to manifest itself similarly to other alien or somewhat alien groups. Although the Maoris are not markedly alien to New Zealanders, this movement has probably been effective. The second factor is related to a temporary maladjustment consequent upon large-scale movements of the Maoris from country to city during the war to meet urgent labour needs. Beaglehole (1955) supplies some figures on this problem.

When one studies the variability of the attitude scores one notices how closely the measures correlate with the attitude scores themselves. The standard deviation values when correlated with the means of the attitude scores, give a coefficient of 0·73 with the German values, 0·67 with the Japanese and 0·63 with the Maori values. In general, therefore, as the attitude scores increase or as the attitudes become less favourable, they become more variable or more heterogeneous. As they become more favourable they become more homogeneous or the situation becomes simpler. The variation for the three racial groups tended to increase with the later years of the war period and for some time after. By 1953 the attitude toward Germans was about as simple or homogeneous as at the beginning of the war, that toward Japanese had not moved completely back and that toward Maoris remained still relatively complex.

III. COMPARISON WITH A STUDY BY REMMERS & WOOD

Remmers & Wood (1947) had been conducting a similar study at about the same time in the United States. Using similar attitude scales they assessed the attitudes toward Germans, Nazis, Japanese and Jews of elementary psychology students at Purdue University. The years of testing were 1935, 1942, 1945 and 1947. The separation of Germans and Nazis is a real one in the United States which has a large German population. Such a contrast could not be well made in New Zealand by a student population.

The attitudes toward Germans were favourable throughout and most favourable in 1947, whereas those toward the Nazis became more and more unfavourable. The attitude toward the Japanese showed a similar trend to that toward the Nazis. By contrast there was no marked change in attitude toward the Jews, so they offered something of the benefit of a control group. The complexity in attitude to this group showed up in high variability throughout, with no improvement in 1947.

Considering the different racial relationships of a New Zealand student group as compared with an American one, the general findings of these two attitude studies are similar.

IV. SUMMARY

1. Tests of racial attitudes using the Thurstone technique with similar samples of students at Auckland University College were carried out with reference to Germans, Japanese and Maoris. The testing was done each year from 1940 to 1947, then in 1950 and in 1953.

2. Changes in attitude scores and in variability such as might be expected during war and succeeding peace were found with reference to Germans and Japanese. The unfavourable attitude toward the Japanese remained, with high variability, even in 1953. The

attitude toward Germans had returned to indifference by 1953, with lowered variability. The attitude toward Maoris changed similarly but at a more favourable level.

3. The Thurstone technique as used here appears to be a sensitive instrument to detect major changes in racial attitude.

REFERENCES

BEAGLEHOLE, E. (1955). The study of social change. *Trans. Roy. Soc. N.Z.* LXXXII, 927–35.

REMMERS, H. H. & WOOD, W. E. (1947). Changes in attitude toward Germans, Japanese, Jews and Nazis. *School and Society*, LXV, 484–7.

THURSTONE, L. L. & CHAVE, E. J. (1929). *The Measurement of Attitude*. University of Chicago Press.

(*Manuscript received* 14 *September* 1954)

NOTES ON THE SHAW TEST

By D. B. BROMLEY

Department of Psychology, University of Liverpool

These notes are intended to publicize the Shaw Test and to suggest the setting up of a Research Exchange through which research work on the test may be facilitated. Full details, photographs, test instructions, etc., can be obtained from the author, and a copy of the test can be borrowed for a short period.

The Shaw Test was first used by Howson (1948). It consists of four wooden blocks which can be arranged in fifteen different series: according to height, weight, colour, position of notch, etc. In addition to these fifteen normal, abstract and clearly defined sequences, which are called Grade 'A' responses, there are at least forty-one less frequent or less clearly defined sequences which are called Grade 'B' responses, e.g. the volume of the blocks, the positional relation between the word and the notch, etc., which are positively but not significantly associated with Grade 'A' responses. There are at least sixteen Grade 'C' responses, mostly illogical, confused and concrete, e.g. making patterns with the blocks, grouping the blocks instead of serializing them, etc., which are negatively but not significantly associated with Grade 'A' responses. Finally, there are at least five Grade 'D' responses, the poorest, mostly failures of comprehension and inane responses, e.g. building towers with the blocks, simply describing the blocks, etc., which are negatively and significantly associated with Grade 'A' responses.

This classification of responses was made through an objective item-analysis of the test performances of 256 relatively normal adults ranging in age from 17 to 82 years. No time limit was imposed and the subjects were, in a sense, 'tested to the limits' of their capacity for producing ideas. The general significance of the responses is that the good and fairly good (A and B) ones indicate the speed, power, originality and general efficiency of the subject's thinking, whilst the fairly poor and poor (C and D) ones seem to indicate the abnormality, defectiveness or degree of deterioration of the subject's thinking. A reliability coefficient for the Shaw Test Grade 'A' responses, computed by the method of rational equivalence, came to 0·87. The correlation between Grade 'A' responses and the Wechsler-Bellevue Efficiency Quotient was +0·745 (+0·641 with age held constant). The correlation between Grade 'D' responses and WBEQ was −0·505 (−0·303 with age held constant).

Other results suggest that the Shaw Test is a promising research instrument for psychometrics and psychodiagnostics. Howson found the test useful in disclosing the fact of but not the degree or location of impairment in brain-injury cases, but there is reason to believe that he did not make full use of the test's potentialities. The safest judgement that can be made at the moment is that the test is a fairly reliable measure of *fluid intellectual ability*. It is one of the few if not the only performance test involving the mental operation of *serializing*, and this fact alone makes the test interesting. If it is used as a timed test on superior subjects, then it almost certainly involves some kind of *perceptual fluency*, and may be of help in defining those factors which refer to high-grade thinking abilities (cf. Guilford, Kettner, and Christensen, 1954). Experimental investigation shows that

Shaw Test performance is influenced by age in exactly the same ways as *creative intellectual achievement* (cf. Lehman, 1953). The test can be adapted as a measure of *originality or conformity*; Klopfer & Kelley (1946) describe a similar classification of responses and give them a significance that could well be applied to Shaw Test responses: popular (frequently occurring) good responses indicate a degree of normal activity and conformity, whereas original (infrequently occurring) good responses indicate an above average degree of activity and brilliance or creativity. It is now commonly understood that tests may not measure the same processes in different subjects or under different conditions, and the Shaw Test is probably no exception.

The fact that Shaw Test performance can be analysed qualitatively and quantitatively in an objective manner, plus the fact that C and D responses are symptomatic of a deteriorated intellect, suggest that the Shaw Test may be useful diagnostically, since abstraction tests are among the more sensitive of instruments for detecting disordered and defective thought processes, and any promising addition to current test-batteries deserves careful consideration (cf. Yates, 1954). An examination of the poor quality (C and D) responses obtained from normal old people suggests that normal intellectual decline approximates to a condition not unlike that found in mental defect and brain injury. A close examination of the number and quality of Shaw Test responses may reward the clinical research worker by helping him to answer questions about and make differential diagnoses in the various psychiatric and neurological disorders, since C and D responses are *signs of impairment*. A Shaw Test Deterioration Quotient can be calculated by taking: $C+D\times 100/A+B+C+D$; this D.Q. has some advantages over that used by Wechsler (1944), but needs further examination.

The main advantages of the Shaw Test are that it is a relatively novel psychological test and adaptable to a variety of uses. It is compact and easy to construct, convenient and quick to administer; the scoring is objective and the task is interesting to the subject. The test's disadvantages are that no norms are yet available, other than those compiled by this author on 256 normal and superior adults, and validity studies are incomplete—but this is unavoidable with new research tools.

REFERENCES

GUILFORD, J. P., KETTNER, N. W., and CHRISTENSEN, P. R. (1954). A factor-analytic study across the domains of reasoning, creativity, and evaluation. *Rep. Psychol. Lab.* no. II. Los Angeles: University of Southern California.

HOWSON, J. D. (1948). Intellectual impairment associated with brain-injured patients as revealed in the Shaw Test of abstract thought. *Canad. J. Psychol.* II, no. 3. 125–33,

KLOPFER, B., and KELLEY, D. M. (1946). *The Rorschach Technique.* Yonkers, U.S.A.: World Book Co.

LEHMAN, H. C. (1953). Age and achievement. *Mem. Amer. Phil. Soc.* XXXIII. Princeton University Press.

WECHSLER, D. (1944). *The Measurement of Adult Intelligence*, 3rd ed. Baltimore: The Williams and Wilkins Co.

YATES, A. J. (1954). The validity of some psychological tests of brain-damage. *Psychol. Bull.* LI, no. 4, 359–79.

(*Manuscript received* 18 *June* 1955)

PUBLICATIONS RECENTLY RECEIVED

Psychology as the Study of Behaviour. An Inaugural Lecture by O. L. ZANGWILL, Professor of Experimental Psychology in the University of Cambridge. Cambridge University Press. 1954. Pp. 36. 2*s.* 6*d.* net.

The appointment of the second professor of Psychology at Cambridge, in succession to Sir Frederic Bartlett, is an event of exceptional interest and importance. As two points are needed to define a line, so presumably a succession of professors is required either to establish a tradition or to define a direction of change.

Cambridge, one might have supposed, was destined from eternity to be the home of a very 'pure' psychology and perhaps of a very distinctive 'school'. In fact it has been the home of neither. The history of its psychological laboratory reflects the influence of two world wars and its most characteristic contributions to science have been in the study of very practical skills. Under the statesmanly direction of Sir Frederic Bartlett it has established a reputation for applied psychology conducted in the spirit of fundamental research. So far from forming a 'school' those who have passed through Cambridge to chairs of psychology and other positions of responsibility are characterized by the diversity of their interests, convictions and talents. Perhaps the only thing they have in common is a comparative immunity from what has been an occupational disease of psychologists, addiction to the membership of schools.

And now we ask: Whither psychology at Cambridge?

This inaugural lecture (delivered on 12 November 1954) is both a review of the history of psychology at Cambridge, and a statement of its new professor's view as to what psychology is and what it should be. The two themes are interwoven since Professor Zangwill, though not overmuch concerned with the applications of psychology, is very much concerned in other ways to preserve and develop the approach of those who have made psychology at Cambridge what it is to-day.

Psychology, no doubt, is what psychologists actually do, and in the end there may be little more that can be said; but the fact that what psychologists actually do in some places is to be amateur philosophers, and at others to be amateur neurologists, naturally engenders a desire for a definition a little less ostensive. Psychology, in this lecture, is defined as the Study of Behaviour. But let no one be mistaken: the Study of Behaviour is not Behaviourism in any of its varieties catalogued up to date. Professor Zangwill picks his way with the utmost circumspection, avoiding the many snares in the path of those who would so describe their subject. He begins by paying due tribute to McDougall, the first to define Psychology in this way, but he shares with other naturalistic psychologists a certain distrust of McDougall's concession to tradition. Whilst deploring Watson's 'brash pronouncements on the non-existence of mind', and regretfully doubting whether advances in the study of animal behaviour have led to corresponding advances in the understanding of human behaviour, he believes, nevertheless, that the attitude of mind which Behaviourism reflected has made possible empirical advances on a broad front. He regards the rapprochement between psychology and physiology as the most important outcome of the Behaviourist movement. If it is for this reason that he gives Watson and his associates very high marks, it is for the same reason that he gives very low marks to the Neo-Behaviourists who in their desire to emancipate psychology not only from philosophy but also from neurology and so to establish a science of 'pure behaviour' have replaced the brain by a tangle of hypothetical constructs and intervening variables.

Professor Zangwill's most positive statement of his own position is framed with characteristic caution: 'I do not regard a human psychology based on consideration of nervous structure and function as wholly without the bounds of possibility.'

How far his own position is removed from doctrinaire Behaviourism is clear when he 'ventures to predict' that at no time will even the sensory physiologist (still less presumably the psychologist) be able wholly to dispense with the testimony of the human observer. If there were a doubt concerning the breadth and tolerance of Professor Zangwill's approach to his subject it would not survive the final statement of his *credo*. This is implicit in a warm appreciation of W. H. R. Rivers: 'I do not believe that we can go far wrong if we model our conception of psychology on the interests and outlook of its first accredited teacher at Cambridge.' This, of course, lets in everything. Rivers was a natural scientist with an exceptionally flexible mind. He could be physiologist, anthropologist, introspectionist or psychoanalyst in his methodology as the problem seemed to require. Where the memory of Rivers is preserved, it would certainly be hard for things to go far wrong. The Department of Psychology at Cambridge is in safe keeping.

C. A. MACE

Experimental Psychology. By B. A. FARRELL (editor) and others. (A Series of Broadcast Talks on Recent Research.) Oxford: Basil Blackwell. 1955. Pp. xi + 66. 7*s.* 6*d.*

This short volume places on record six broadcast talks given by different speakers in the Third Programme of the B.B.C. in the summer of 1954. The common bond between the speakers lies in their connexion with the development of psychology in the Universities of Oxford, Cambridge and Reading. The individual topics belong to a very wide conception of experimental psychology. The sample covered included: I. Perception (Watson); II. Adult Learning and Remembering (Kay); III. Motivation (Deutsch); IV. Some Hypotheses in Psychoanalysis (Farrell); V. The Study of Social Behaviour (Argyle); VI. The Prospects of Experimental Psychology (Oldfield). Each contributor dealt with a topic of special personal interest.

As the editor correctly indicates, a different group of speakers would have chosen a different set of topics. Furthermore, no claim is made that these talks are to be considered as contributions to psychology. They are directed towards the intelligent and educated layman in an attempt to remove certain current 'misconceptions which surround the work of the experimental psychologist'. From that standpoint the book is of considerable interest to all who are concerned with the 'public relations' of contemporary psychology. The 'recent research' referred to covers roughly the last fifteen years. Only in one talk (II) is there substantial emphasis on work carried out by the particular contributor. In most cases the talks are topical reviews in a conversational style of important literature. In spite of an excellent final talk by Prof. Oldfield giving a perspective to the whole series one wonders whether the layman will not feel that the contributions of experimental psychology are rather trivial. Is such a distortion an inherent limitation of a short exposition on a complex series of topics? Does the popularization of psychology demand some form of anecdotal treatment of a type from which academic psychology has only recently emancipated itself? Such questions cannot be examined here.

From the point of view of the serious student of psychology this book might well be recommended as background reading. Its broad perspective should be valuable to the student in the early stages of courses in experimental psychology when motivation tends to decline owing to the necessarily segmental and artificial nature of elementary experimental work.

ROLAND HARPER

Vision with Spatial Inversion. By F. W. SNYDER and N. H. PRONKO. Wichita, U.S.A.: University of Wichita Press. 1952. Pp. 144. No price given.

The authors describe an experiment in which the subject wore inverting spectacles for 30 days, covering the eyes with a bandage whenever the spectacles were removed. The spectacles were in the form of 'unit-power' inverting telescopes, attached by appropriate head-gear. They gave a clear binocular field of 20 degrees diameter, and were set in a slightly convergent position in order to permit optimum binocular vision for near objects. The subject carried out systematic practice without the inverting lenses on the following tasks in the 'pre-inversion' period: card sorting, Purdue Pegboard, Minnesota Rate of Multiplication Test, mirror tracing. Then he continued the learning tasks with the inverting spectacles in place, for 30 days, and finally completed the practice at the tasks after the spectacles were permanently removed. Full details of the learning experiments are given. The narrowing of the binocular field to 20 degrees alone did not interfere to any extent with the tasks, but when the lenses were first put on the visuo-motor co-ordinations were seriously disrupted. Later the subject gradually adapted himself to the inverted vision and improved at the learning tasks even more than before the lenses were worn. Removal of the lenses again disrupted the visuo-motor co-ordinations for a short time. The authors infer that space perception does not exist at birth, but comes into being as a result of a series of 'contacts' between the organism and the environment. The subject's own observations make it abundantly clear that the question whether there was a true phenomenal reversion of the inverted image to uprightness would have been a false problem. At first he saw objects upside-down and inverted from side-to-side. Later, when visuo-motor co-ordinations were fully established he was able to read, sketch, to walk about and even to drive a car, and it did not occur to him to doubt the uprightness of his visual pictures. When asked whether things looked upside-down, he replied, 'I wish you hadn't asked me. Things were all right until you popped the question at me. Now, when I recall how they *did* look *before* I put on these lenses, I must answer that they do look upside-down *now*. But until the moment that you asked me I was absolutely unaware of it.' What he saw had appeared quite natural until his companion asked him, just as much so as in his pre-inversion days. When a clown was seen to jump off the diving board into a swimming pool he appeared to move downward into the pool, but in fact the way he seemed to move depended upon whether or not the subject made reference to his memory of the pre-inversion way of seeing things.

R. W. PICKFORD

Destiny and Motivation in Language. By A. A. Roback. Cambridge, Mass.: Sci-Art Publishers. 1954. Pp. 474. $8.50.

The subtitle of this volume is *Studies in Psycholinguistics and Glossodynamics*. In a copious Glossary which is—fortunately—provided 'Psycholinguistics' is defined as 'The phase of linguistics which introduces psychological considerations'; and 'Glossodynamics' as 'The psychological phase of linguistics which deals particularly with motivation either on the sound level, meaning level, or grammar level'. It is to be doubted whether these definitions make anything very much more clear. Psychological considerations seem as if they can be introduced into *any* phase of linguistics, while the motivation that is referred to might be whatever active human processes can be supposed to give rise to particular sounds, meanings or grammatical forms and their derivations over a long period of time; or it might be whatever activities particular sounds, meanings or grammatical forms themselves set up, or it might be both of these, or, apparently, it might be neither of these, but something else. The fact is that this is a book into which Dr Roback can introduce anything whatever that he wants to say about language. In the result the volume is an excessively difficult one to read straight through, for it turns and twists about all over the place according to what may perhaps be called the very lively motivation of its author. But the psychologically minded reader who dips, and then comes to the surface and ponders a while, and later dips again, will not only get a lot of enjoyment out of the exercise, but will also probably be started upon many interesting and profitable explorations.

Part I of the book deals with 'historical symbolism', and this is where most of the valuable and unconventional discussion about 'destiny' is to be found. Part II is headed 'Glossodynamics', and, broadly, is concerned with psychological forces that may have led to language, to special languages, and to the interplay of languages. Part III is called 'ethnolinguistics', and a good many readers are likely to find this the most attractive and original part of the whole treatment. It contains many significant and often brilliant comments upon language forms and constructions that have become particularly frequent or favoured in various social and cultural groups.

Nobody could reasonably call *Destiny and Motivation in Language* an easy work to read. Its range of learning is prodigious, its style of writing is often a mixture of the severe and proper with the airily conversational, and the direction of the argument seems sometimes to get rather lost among ramifications of detail. But it is a book which nobody who is, or wants to be, interested in the psychology of communication can afford to neglect, though it has no truck whatsoever with Cybernetics or Information Theory. Once more, Dr Roback has shown that it is extremely difficult for him to write at length about anything without being profoundly interesting.

F. C. BARTLETT

Child Psychology. (Fourth edition.) By Arthur T. Jersild. London: Staples Press. 1955. Pp. xii + 676. 42*s*.

This most recent edition of a well-known text on Child Psychology is described as 'virtually a new book'. The previous editions have been brought up to date by the inclusion of recent material, but the basic plan and approach is still recognizably that of the earlier versions. The emphasis is still on the total growth of the child, with particular attention being given to the emotions and personality. There is no doubt that the revision has been an improvement; the particular virtue of the latest edition being not so much in the inclusion of recent studies, as in the deeper and more mature treatment of children's development. There is much wisdom and humanity in it, but not at the expense of science. The excellent bibliography deserves special mention.

Those who are familiar with the earlier versions will perhaps not find as great a change as is claimed; to those who have not made their acquaintance with this book, it can be highly recommended.

JAMES MAXWELL

The Study of Personality: a Book of Readings. By HOWARD BRAND (ed.). New York: Wiley and Sons; London: Chapman and Hall. 1954. Pp. xvi+581. 48*s.*

This book attempts a cross-sectional view of research on personality. The editor has written introductory passages and there are thirty-one contributions, principally from journal articles. Contributions in the 'Theory' section include articles by Alexander, Allport, Eysenck and Sullivan; in the 'Methods' section are Stephenson on Q-technique, Thouless on psychopathology, and others; finally in 'Problems' Newcomb writes on role behaviour, Hilgard and Maslow on motivation, while Sears, Frenkel-Brunwik and others are also represented. Yet despite these and other distinguished contributors the book is a disappointing one; its weaknesses are of editorship in that the articles have been badly chosen. To illustrate, from G. W. Allport's large bibliography two papers are published. The first comprises Allport's unsympathetic assessment of Stern who, one feels, might well have been permitted to provide his own exposition; the second article consists largely of Allport's comments on Mr Bertocci's opinion of Allport's earlier comments on McDougall! For a book of this kind many of the other papers included would seem equally difficult to justify. Again, some of the strongest parts of the book—N. E. Miller on displacement and Eysenck on the literature of schizothymia—are readily available in quite recent numbers of the principal APA or other well-known journals.

As a whole this book does not achieve its stated aim, and moreover certain individual contributions merit rigorous criticisms in terms of their scientific worth. Thus C. S. Hall deals with dream analysis from a standpoint which adds nothing important to the theory of Rivers, and proceeds without either evidence of awareness of, or influence by, this earlier and better worked-out formulation. Again, Farrell's chapter on the scientific testing of psychoanalytic work proceeds with the apparent assumption that such 'testing' can be achieved largely by citing the names and dates of relevant investigators. The argument thus comes to depend more on the prestige of names than on any attempt at evaluation of the scientific quality of the investigations themselves: surely a poor lesson in scientific methodology. Finally, a major defect is that two of the most vigorous and important developments in personality study—Learning Theory and the Functional Perception movement—are both badly under-represented.

PETER McKELLAR

The Interpretation of Dreams. By SIGMUND FREUD. Translated and edited by JAMES STRACHEY. London: Allen and Unwin. 1955. Pp. xxxii+692. 21*s.* net.

As the translator points out, Freud regarded *Die Traumdeutung* as his most important work. 'Insight such as this', he wrote, 'falls to man's lot but once in a life-time.' The edition here reviewed is practically identical with that contained in vols. IV and V of the standard edition. Since these volumes are not obtainable separately from the set, this edition is assured of its welcome. In addition to being an entirely new translation, it is a Variorum edition, with full indications of the stages at which Freud introduced his many alterations in successive versions of the text; the supplementary material—bibliographies and indices—has been greatly expanded and completely recast.

Strachey's translation is perhaps freer than Brill's, and his style probably more remote from Freud's own, but greater clarity has usually been achieved. Most compound German nouns (of which *Traumdeutung* itself may be regarded as a prototype), customarily represented in translation by hyphenated equivalents, are now rendered into more idiomatic English usage. Many other changes in minor but familiar detail are noticeable: e.g. *secondary elaboration*—regarded as a mistranslation—is replaced by *secondary revision*. On the other hand, the word play so characteristic of the dreams recounted by Freud is dealt with very literally, even pedantically, a good fault—if it is a fault at all—in a work designed to be scientific rather than entertaining.

The book is admirably produced, and most inexpensive by present-day standards.

BORIS SEMEONOFF

The Mythology of the Soul. By H. G. BAYNES. London: Routledge and Kegan Paul, 1955. Pp. xiii+939. 50*s.* net.

The re-issue of this important book will be acclaimed by analytical psychologists and psychiatrists. C. G. Jung's monumental contribution to the modern methods of psychotherapy and his immeasurable influence on Western thought and feeling unceasingly call for new evaluations and elaborations of case material by those analysts who—like Baynes—are using the concepts of the Collective Unconscious (i.e. the impersonal unconscious of mankind, the race, the nation, the family) in their work and thought. As at first most of the analytical literature naturally stemmed from Middle European sources, it was fitting that Baynes should be one of the first contributors in Britain. Himself a pupil and friend of Jung, he remained for many years until his premature death in 1944 the inspired leader of the Jungian School in England, and his original and congenial personality has left its distinct mark on Analytical Psychology as it has grown from its own collective roots in these islands.

Baynes was, indeed, undeniably British. A foreign admirer once whispered at the onset of one of his lectures: 'This man is surely not a psychologist: he is an English Gentleman.' The reader of his *Mythology of the Soul*, however, will be equally impressed by the depth of his psychological knowledge and his constructive intuition as by the width of his cultural background and the sureness of touch in his investigation.

The task which he sets himself in this scholarly work is the detailed evaluation of analytical material of two schizophrenic patients. Schizophrenia is perhaps *the* mental illness most typical of man's inherent psychic condition. Religious and philosophical systems have recognized the existing split since the beginning of time, and while the former have allowed mythological expression of it, they both have sought to devise formulae for the solution of the inner conflict in mankind's divided soul. Baynes has no formula to offer and while he uses the relevant mythology to the fullest possible extent, he endeavours to reach down to the very roots of the catastrophic split which we all have to accept as our own to some degree in order to kindle the therapeutic light in the darkness within and thus begin to heal the great ill. He uses the individual material to elucidate man's fundamental problem, and his readers can hardly fail to follow his inspiring leadership into their own depth.

The book is full of valuable and most original teaching. One may at times feel inclined to separate some of his revealing formulations and descriptions of psychological conditions from the drama of the individual dreams and drawings, because the latter is naturally far less accessible in all its microscopic detail to the reader. He must lack the transference link which exists between the analyst and his patient, in spite of Baynes's warm-hearted and capable handling. The derived formulations, on the other hand, which are at times threatened to be drowned by the enormous power of the analytical drama, could well fill a handsome volume of psychological aphorisms of great and lasting value.

The book is beautifully produced, and the publishers must be congratulated on the skilful reproduction of the many drawings. To facilitate the study of the invaluable material, it would still be better if the drawings could be on separate sheets loosely attached to the volume so that the reader could compare text and material at greater ease.

W. P. KRAEMER

The Many Faces of Love. By HUBERT BENOIT. London: Routledge and Kegan Paul. 1955. Pp. vii+308. 28*s.* net.

This book, translated from the French by P. Mairet, is written by an ex-surgeon who, after being injured in the War, turned to the study of psychology. Dr Benoit has clearly a most sensitive appreciation of human nature coupled with a love of philosophy—indeed his work might appeal more to the philosopher than to the scientist. He considers that there are three basic aspects of love—adoration, benevolent and appetitive love, and he investigates the ways in which they fuse and interact. He also discusses various perversions and inhibitions of normal love, but although he has implicitly taken over some of the concepts of psychoanalysis, his thoughts will sound foreign to those who believe in the importance of the Oedipus complex, castration anxiety, and early object-relations in the development of the ability to love.

The work itself is written in the form of a dialogue between the author, a young man and a young woman. In spite of this literary form, it is extremely difficult to read. Perhaps this is best illustrated by a quotation:

'*The Author*:...In the last stage of man's realization, the realization that is timeless, the triangle is constituted otherwise: the passive, feminine pole is represented by the totality of the temporal aspect of the man, both his "soma" and his "psyche"; the masculine, creative pole is the Creative Principle which is at the centre of every created being, the nucleus we sometimes call "Spirit", in the sense in which that "Spirit" is prior to both the "body" and the "soul". The "Spirit" penetrates the psychosomatic organism of man, in an ultimate interior "coitus", by virtue of a hypostasis which is then the very principle of Love, for which the scholastic term is "Agape", or the active love of the Supreme Principle for his passive creature. The result of that act of love is, then, the "new man", man "liberated", "realized", reunited to the Principle of the whole of manifest being.'

'*The Young Man*: I confess, sir, that these views bewilder me.'

The reviewer agrees.

JOSEPH SANDLER

Social Psychiatry. By Maxwell Jones. London: Tavistock Publications Ltd. 1952. Pp. xix + 186. 18*s*. net.

'The development of community methods of treatment implies the wider application of the social sciences, mainly sociology and anthropology to existing psychiatric practice', writes Dr Jones in his introduction. Along with chapters written by his colleagues he reports on some five years' work with a therapeutic community, the Industrial Neurosis Unit at Belmont Hospital, Sutton, Surrey. Two early experimental communities, the Effort Syndrome Unit at Mill Hill Emergency Hospital, and the Ex-Prisoner-of-War Unit at the Southern General Hospital, Dartford, are described. By May 1945, when the latter unit was begun, three concepts were emerging: treatment was a continuous process operating throughout the entire waking life of the patient in hospital: a re-organization of the hospital society was needed with a greater degree of social penetration between patients, nurses and doctors: the patient's role in the hospital society needed scrutiny.

The third community, with which the remaining chapters are concerned, was established in April 1947 at Belmont Hospital with the backing of the Ministries of Health, Labour and Pensions. Its main function was to study the chronic unemployed neurotic. The population of the unit (100) comprises in the main chronic neurotics and character disorders of a kind usually considered unsuited for treatment by psychotherapy or physical methods.

Dr Jones defines a role as a set of legitimate expectations of behaviour and then proceeds to outline the social structure of the industrial neurosis unit with nurses (authoritative social therapeutic roles), doctors (social supportive exemplary activating and interpretative), disablement resettlement officers, social workers, psychologist and patients.

The three following chapters by Dr Jones, by Dr Thomas Freeman and by Dr B. A. Pomryn offer a more detailed account under the heading of social therapies, in-patient psychotherapy and techniques of group formation. The viewpoint here is non-specificist. Neurosis is not an illness—'by neurotic illness we mean emotional difficulties, conscious or unconscious, which result in an inability on the part of the individual to meet the needs of his environment, or to make adjustments (at a conscious or unconscious level) sufficient to maintain an adequate functional efficiency'. How such a group of persons, staff and patients, function therapeutically is illustrated by an account of the unit discussion groups and psychodrama as part of the subculture of the unit.

Six months after the first industrial case was discharged from hospital, a follow-up study was begun. Two groups were investigated: (*a*) patients who had attended the industrial neurosis unit, and (*b*) disabled neurotic unemployed people, who had not attended a rehabilitation treatment centre such as the Unit, and who were referred by the Employment Exchanges of the Ministry of Labour. 104 of 238 consecutive discharges were included. Joy Tuxford outlines the follow-up interview and the general aspects. In the course of this chapter there are several sociological impressions for which one hopes someone else will provide evidence—'there was in most areas a real paucity of group life', or 'the family, which was in most instances a small and loosely knit group, was often the only one which had any real meaning for the individual'.

Joseph Sandler writes on the statistical analysis of the follow up and on the concept of general adjustment. An adjustment score was calculated for each patient and related to the information recorded in hospital. 'The fact that a number of antecedent measures are significantly related to adjustment means that it is possible to predict which patients will do well and which badly, to a degree of accuracy better than might be expected by chance alone.' Unit doctor's diagnosis, doctor's assessments of improvement in hospital and marriage, for example, are associated significantly with good adjustment. Six months after leaving hospital two-thirds of patients followed up had made a fair adjustment or better: just over one-half had worked full-time since hospital.

Doctor Sandler's chapter on Vocational Guidance contains some very important comments and is recommended not only to the vocational guidance psychologist but to the clinical, educational and industrial psychologist.

Dr Jones writes a final chapter: 'Our findings appear to justify the conclusion that it is possible to change social attitudes in relatively de-socialised patients with severe character disorders, provided they are treated together in a therapeutic community.' He goes on to talk of the hospital in the light of some of the concepts of social science which will be familiar to most readers of this *Journal* but are likely to be rather revolutionary to some medical men.

There is some compensation for the lateness of this review in being able to recommend also the article by Baker, Jones, Pomryn and Merry, 'A Community Method of Psychotherapy' in the *British Journal of Medical Psychology*, XXVI, 222–44, 1953, which is, as the authors say, 'complementary to our recently published book'.

For some readers the term Social Psychiatry is likely to excite as much emotion as Social Medicine, but names are not all-important. This book is recommended to both doctor and psychologist. It overstresses neither heredity nor environment: it sees behaviour as a resultant, and modifiable. It is a modest account of an exciting experiment.

JOHN C. KENNA

Einführung in die medizinische Psychologie. By GEORG DESTUNIS. Berlin: Walter de Gruyter and Co. 1955. Pp. xi+218. DM. 22.

Half of this book is devoted to the normal and the other half to the neurotic personality. 'Personality' is defined exlusively in terms of philosophical psychology, and formulations of the Leipzig *Ganzheitsschule* (Krüger-Sander-Wellek) are mainly preferred. Similarly, intuitive assumptions on the nature of traits, their integration and disintegration, are accepted throughout in preference to operational definitions.

Normal personality, in the author's view, is made up of fourteen drives of maintenance, sex, aggression, curiosity, etc. In addition, consideration is given to motivation, sentiment, affection and emotion and, finally, the six noetic functions, will, perception, imagery, memory, thinking and intelligence. A remark directed against the definition of intelligence by Spearman and others that 'these concepts were only considering the elements (i.e. neglecting the whole)' expresses the author's desire to decline any factual analysis.

In the following chapter on psychodiagnostics a few known tests of intelligence, child development, association and projection are briefly and uncritically reported. A description of the historical continental typologies concludes the first part.

Part II, on neurosis, presents in an elementary fashion a comprehensive variety of present-day German theories, and then deals with conventional descriptions of neurotic psychogenic, psychosomatic and psycho-organic symptomatology. A short chapter on neurotic-like reactions during illness is added.

The bibliography consists of almost 250 items, roughly 85% in German, 10% in English and 5% in French. Half of these date back 20 years or more. Considering the wide range of topics covered in this book, this selection of references suggests that the author's attempt 'to summarize and to evaluate the most recent results in the field of medical psychology' can hardly be considered to have been achieved.

J. C. BRENGELMANN

La notion de normal en Psychologie clinique. By FR. DUYCKAERTS. Paris: Libraire Philosophique J. Vrin. 1954. Pp. 206. No price given.

This book, published in a series called 'Problèmes et Controverses', is a philosophical discussion of the psychological notion of normal behaviour.

The author opposes those who contend that normality can only be a relativistic notion, either because they fear the contamination of psychological science by ethical judgements or because they are struck by the differences in accepted behaviour of separate social groups. He attempts to build up an objective definition of normality and the use of psychotherapy is then for him *ipso facto* justified.

The author chooses to discuss normality first as 'integration', then as 'autonomy' in relation to other individuals, as 'adaptation' to different social groups, and finally as resemblance to the average. The consideration of normal behaviour as 'integrated' behaviour is by far the most interesting and stimulating chapter. The final definition of 'integrated' behaviour is related to observations made by such authors as Schwarz, Goldstein and Maier. The common thesis of these authors is to distinguish between the significant, ordered, motivated forms of behaviour on the one hand, and the catastrophic or frustrated forms of behaviour on the other, and on this distinction in the main the author bases his definition: to be normal, behaviour must be purposeful.

Autonomy and adaptation are perceived as complementary to each other, and from an analysis of their meaning, the author derives an added necessary qualification of his definition of normal behaviour; it must be creative, in other words, it must include an assimilation of the past and of surrounding cultural values and aim at the building up of effective social contact.

The concept of normality identified with statistical average is rejected with horror. Kant is called in to help in the forty-page demonstration that such heresy does not fulfil any of the criteria of a logical definition, and in a mood of angry humour the author attempts to unveil the motives of those whom he believes to uphold such a notion.

The book will disconcert the reader who is more used to Anglo-Saxon literature. It is so frankly and unashamedly theoretical and philosophical, so devoid of any empiricism, so direct in utilizing facts as examples illustrating the author's ideas—instead of building those ideas on a succession of facts.

In the tradition of French contributions to philosophy and to science, the weightiest part of the book is the critical discussion and clarification of ambiguous notions, but when the author comes to the construction of practical concepts, he is guilty of certain evasiveness. The norm which the author finally defines is both absolute and yet infinitely varied in its mode of expression, but although it thus satisfies not only logic but also elegance of thought, it defies any attempt to make it serve as an operational and useful criterion.

MARYSE METCALFE

Psychoanalytic Interpretation in Rorschach Testing. By ROY SCHAFER, PH.D. Pp. 446. New York: Grune and Stratton. $8.75.

In 1923, with the posthumous publication of the case-study entitled 'The Application of the Interpretation of Form to Psychoanalysis', the marriage of the Rorschach method with psychoanalysis was announced by its founder. Dr Roy Schafer cements this fruitful alliance in his valuable book, which to a notable extent co-ordinates and enlarges the work of other writers in the field, adding much that is original with a refreshing clarity of reasoning and economy of words. As is stated in the preface, this is not an easy book to read, for to profit by it adequately there is required not only a thorough grounding in Rorschach principles and technique but also considerable familiarity with psychoanalytic literature and methods. This warning, however, should not deter any serious Rorschach worker, since at every point hypotheses are clearly stated before any conclusion is drawn, so that the reader is gently educated as he goes.

Main themes discussed are:

(*a*) Interpersonal dynamics in the test situation.
(*b*) The response process.
(*c*) Thematic analysis.
(*d*) Criteria for judging the adequacy of interpretations.
(*e*) The interpretations of defence-mechanisms.

In recent years there has been an increasingly healthy recognition that the personality of the tester, and especially his unconscious needs, may seriously influence test results as well as their interpretation. The 'toilet of the analyst' is no less important for validity than is the intra-personal discipline of the clinical psychologist engaged in exploring the personality of others. Without such discipline he may lay undue stress on the voyeuristic aspect of his role, or, feeling anxious or guilty about this, may swing to an attitude described by Dr Schafer as that of a psychological ringmaster, his insistent directions arousing in the patient 'anxiety, inadequacy, compliance, rebellion, and a host of other reaction tendencies'. Other unconscious drives may cause the tester to stress the magically omniscient, autocratic, aspect of his role or the 'saintly', also with consequent effects upon the subject. Appropriately, this discussion of the dynamics within the tester is followed by an equally careful study of the dynamics of being tested. Dr Schafer never forgets the patient in a test situation and shows a rare understanding of the feelings which this may arouse.

In the chapter on 'The Response Process', some of the test instructions, based on those of Rapaport, might be challenged as over-directive or seductive, but the description of shifts in the level of psychic functioning, associated with Freud's concepts of 'primary process' and 'secondary process' thinking, is both valuable and illuminating. Both processes may be observed in most Rorschach records, for although 'on the advanced levels of psychic functioning the reality principle and the secondary process prevail over the pleasure principle and the primary process'...'the primitive levels do not go out of existence.... They frequently co-determine responses with the advanced levels, putting a personal stamp—possibly too personal—on would-be impersonal, objective reactions.' In this way there is demonstrated the interplay of the forces of id, superego and ego which produce a picture of the man as he is with his child-self within him, integrated or unintegrated. The value for diagnosis of this assessment of dynamics is obvious, as is the later study of defences.

The chapter on thematic analysis, though partly schematic, includes an original contribution to the dynamics of content analysis, its uses and pitfalls, these latter being emphasized in the following chapter. All those using projective tests, especially those who tend to jump to conclusions when interpreting Rorschach material, will profit by studying Dr Schafer's criteria for judging the adequacy of interpretations. These provide a salutary check upon intuitive and get-there-quick methods with over-facile deductions.

The long section on the interpretation of defences, which comprises two-thirds of the book, is of outstanding value. Others with psychoanalytic orientation have taught that analysis of the mechanisms of ego-defence is essential for full understanding of Rorschach data, but here there is published for the first time a well-documented and expanded study of the subject with a wealth of illustration to clarify its points. From an admirable summary of the psychoanalytic concept of defence the reader is led to consider defensive operations in the Rorschach response process. In his introduction to this section and in the subsequent chapters the author uses all that he and the Rorschach method have got, and the result is impressive. 'Defensive operations' are related to scores, to images and attitudes, and to all combinations of these. Various mechanisms of defence are discussed in great detail as illustrated by psychogram and sequence analysis combined, demonstrating such mechanisms as repression, denial, projection, and the obsessive-compulsive. Although others might have been included, those quoted serve to demonstrate the function of unconscious defence in the healthily adjusted, as well as in the fields of neurosis and psychosis of varying types.

The level of this publication is so high that one is led to wish for expansion in some respects, as, for example, a more sensitive scoring of shading that would illuminate the tactile responses as associated with early feelings in the mother-child relationship. More about depression might also have been welcome, but it is unfair to ask for everything, and the reader can certainly give thanks for what he has received, with respectful congratulations to the writer.

THEODORA ALCOCK

The Adult-Child Interaction Test: a Projective Test for use in Research. By THERON ALEXANDER. University of Illinois, U.S.A. Monograph of the Society for Research in Child Development Inc., Vol. XVII, Serial no. 55, no. 2. 1952. Pp. v+40, with 8 plates. No price given.

'This test is planned to give information primarily about perceptive experience of adults in reference to children and children's perception of adults.' As such the test contains the germ of a valuable idea for a projective approach. But the method here developed is disappointing. The rationale offered neglects the more important questions concerning perception and personality, nor does it give any framework within which to think about motivations in terms of interpersonal relations. In the few pages devoted to the theoretical basis of the test we find vague generalizations which gloss over these important issues. 'Broadly speaking, human behaviour consists of the selection of stimuli for admission to the organism (controlled by the use of energy and reactivity of the receptors), the assignment of meaning of stimuli coming from the receptors, and selection of responses and the activities of the body system.... Finally, as a result of the organism's experience, characteristic individual approaches to problems and individual perceptions develop.'

With such a general basis and with only the most slender references to any clinical appraisal of behaviour, it is not surprising that the selection of picture situations shows no originality, that the method of analysis provides little of insight into personality dynamics and that the conclusions seem naïve and superficial.

H. PHILLIPSON

The Language of Dynamic Psychology as related to Motivation Research. By JOSEPH W. WULFLECK and EDWARD M. BENNETT. New York, Toronto and London: McGraw-Hill. 1954. Pp. 111. 30*s.*

Sponsored by the Advertising Research Foundation, this well-produced but expensive little book is intended to 'make the books, reports, and writings of dynamic psychologists and motivation research specialists of much more use to advertising men and other readers concerned with human behaviour'.

Taken at its own valuation it affords quite stimulating and at times even entertaining reading. About 600 terms, presented alphabetically, are defined, with illustrative comments. Many psychological and allied fields are represented—even some Rorschach symbols are included. There are, however, some surprising statements; e.g. of *statistical trend*: 'The phrase may be used in much the same way as the word *correlation*.' And under the heading *voyeur*: 'The desire to look at things which are taboo, concealed, and hidden, is trained into little boys at a tender age by their fathers.' BORIS SEMEONOFF

Psychological Tests for Accident Proneness and Industrial Proficiency. By E. G. CHAMBERS. London: H.M. Stationery Office. Medical Research Council Memorandum no. 31. 1955. Pp. iii + 30. 2*s*. net.

This Memorandum summarizes five I.H.R.B. reports by E. Farmer and the author, published between 1926 and 1940, and all now out of print. It contains no new material nor fresh treatment of the data. Over 4000 industrial and transport workers, including Services personnel, were subjected to a wide variety of psychological tests, and accident records for many more were examined. Two main conclusions emerged: that accident proneness was a relatively stable personality characteristic, and that a selective process that increased industrial proficiency would at the same time tend to decrease accident rate. Relationships between accident incidence and many other variables are reported, mainly in terms of weighted mean correlation coefficients; a small selection of tables and graphs from the original Reports is included. The pre-war work is regarded as unfinished, and as now in some ways dated; the author expresses the hope that others will wish to continue this line of research. BORIS SEMEONOFF

English Social Differences. By T. H. PEAR. George Allen and Unwin. 18*s*.

Prof. Pear's interest in the contemporary scene is well known, and this work is an entertaining collection of his thoughts and observations upon class differences in eating habits, in clothes, in speech, in etiquette and a number of other spheres. Prof. Pear has been taken to task by some reviewers because his treatment of the subject is not sufficiently empirical, but his career in psychology, long as it happily has been, would have been far too short for all the topics in this book to be treated, in the fullest sense of the word, psychologically. Let us rather be thankful for a psychologist who is wise enough to draw our attention to problems which he does not himself pretend to have solved, and who is humane enough to do so in a manner which is highly diverting. A. C. T. W. CURLE

American Indian and White Children: a Sociopsychological Investigation. By ROBERT J. HAVIGHURST and BERNICE L. NEUGARTEN. Chicago: University Press; Cambridge University Press. 1955. Pp. xiii + 335. 37*s*. 6*d*. net.

This volume overflows with tables, giving in meticulous detail the results of an extensive series of tests administered to children of six Indian tribes and to a control group in a midwest city. Building upon the work of Piaget, new tests were worked out to study the moral and emotional development of children in such a way as to be applicable in cross-cultural investigations.

Two of the tests proved particularly useful and are commended by the authors to anthropologists as a systematic interviewing technique. In the Emotional Response Test children were asked to describe situations in which they were happy, sad, afraid, angry, ashamed, and to say what were the best and worst things that could happen to them. In the Moral Ideology Test children were asked to name specific forms of behaviour which they considered good or bad, and to name the source of authority for each. The recorded differences between the various Indians, whites, and New Zealand children (separately reported by Havighurst in an appendix) are in no way startling, but the results are such as to permit the authors to make general statements about the children of these communities. MICHAEL BANTON

Growing up in the City. By JOHN BARRON MAYS. Liverpool: University Press. 1955. Pp. lx + 216. 17*s*. 6*d*.

This is a first-class study of 'normal' delinquency, that is delinquency socially determined, connoting no personal instability in the delinquent. It brings a welcome new emphasis into a field overmuch given, in recent years, to equating delinquency with psychologically centred maladjustment. The eighty boys whose experience Mr Mays records are almost without exception healthy boys who have grown up in a highly deficient and delinquent neighbourhood, and who have conformed to its patterns of behaviour. On a conservative estimate sixty-two were actually delinquent and most of the others probably so. Yet for all of them, with one possible exception, delinquency was or looked like being a passing phase, a kind of mild social measles. They had no feelings of guilt. They enjoyed the fruits of theft. They found it exciting—'it makes you sweat', as one boy said. There is a strong element of bravado. Once begun 'you have to keep going to prove you haven't got cold feet'. The main driving force is the impetus and approval of the gang, the small group of pals who go about together. Not all of the gangs are delinquent, but most of them are. They simply perpetuate the norms maintained by their predecessors, now their adult contemporaries. It is remarkable in the circumstances that a large number of boys avoid the more serious types of crime (petty shop-lifting from big stores and lorry-skipping are the more usual practices) and that most of them give it up between 13 and 16. It shows that the new non-delinquent phase, like its predecessor, is simply part of the process of growing up in that particular district. It shows also that even in a 'delinquent neighbourhood' the pressures to conform to the standards of the wider society are strong. The gang often keep together, but as they grow older they become more able to understand the risks they have been running, more reluctant to 'upset Mum', more capable of appreciating the feelings of the injured party.

One critical note. The author under-writes the view that young 'delinquents'—mild and transitory cases of the type he has observed—are primarily 'environmental'; while young 'criminals'—the more serious and incorrigible cases who constitute the hard core of the problem—are victims primarily of severe emotional disorder (caused usually by deficient parent/child relationships). This view is widely held. But the field of delinquency, or rather of the social and personal problems to which delinquency is a pointer, is vast and largely unexplored. Intensive investigation in this country has so far been conducted mainly by psychiatrists and clinical psychologists, who have naturally concentrated on personal disturbance. The further development of direct sociological and social-psychological inquiry on the lines initiated by Mr Mays may reveal a further type of hardened offender who is psychologically healthy; the 'social, gregarious, active and athletic boy', described by the late E. H. Sutherland, who takes on the impress of his delinquent neighbourhood more firmly than those boys who co-operated with Mr Mays to make this book. There is certainly room for inquiry, for example, into the composition of the strong informal groups to be found in a few approved schools and in many Borstals who effectively prohibit fraternization with staff. They are held together, it may be surmised, by leaders who have gone wrong by social rather than by psychological standards.

J. A. MACK

Oppression. A study in Social and Criminal Psychology. By TADEUSZ GRYGIER, with a foreword by HERMANN MANNHEIM. London: Routledge and Kegan Paul Ltd. 1954. Pp. xiv + 362. 28*s*. net.

This is an account of an inquiry into the influence of oppression 'on human behaviour and on aggressive tendencies in particular'. Oppression is defined as a combination of 'five conditioning factors': lack of freedom, disregard of personality, isolation, insecurity, and frustration. In order to show that more intense oppression produces a greater tendency 'to direct aggression outwards', Dr Grygier used some of the T.A.T. material and Rosenzweig's Picture Frustration Study with groups of Polish Displaced Persons in the UNRRA camps in Germany. The groups were very carefully equated in all respects apart from the variable of 'background of oppression': former inmates of concentration camps were compared with former deportees to German industry and agriculture. A subsidiary investigation deals with delinquent Displaced Persons. The most striking finding here is the very high relative proportion of delinquents with the background of concentration camps.

This is an original, painstaking and courageous investigation, clearly written and fascinating. One may regret, however, that more evidence is not provided to justify the author's treatment of the population of the camps as 'homogeneous' at the time of his inquiry, apart from differences of background. This is possibly related to Dr Grygier's deep commitment to the 'criminal personality' view of delinquency. As he puts it himself, statements to the effect that delinquency may be an outcome of social norms and values 'explicitly contradict the whole basis of our study—the concept of antisocial personality, and the view, hitherto accepted, that the criminal personality breaks the rules imposed by any society or group'.

Dr Grygier's various statements about national character—Polish, Jewish, Russian or German—are on a level very different from the careful discussion of his equated samples of subjects. Here, as in the final chapter on the implications of the findings, the text sometimes tends to be an expression of

Dr Grygier's *Weltanschauung*, psychological and otherwise, rather than a contribution to knowledge. It is difficult to see how some of the author's generalizations on war and international co-operation, human nature, Apollonian and Dionysian societies, etc. are implied by, or even at times related to, the results of his research.

It should be stressed, however, that this does not detract from the value of the research itself, to which the major part of the book is devoted, along with a careful discussion and evaluation of all premises and conclusions.

H. TAJFEL

Integration of Religion and Psychiatry. By W. EARL BIDDLE. New York: Macmillan. 1955. Pp. xii + 171. $3.75.

Psychology and Worship. By R. S. LEE. London: S.C.M. Press. 1955. Pp. 110. 7*s*. 6*d*.

Books on the relationship between psychology and religion are greatly needed, and the books by Dr Biddle and Dr Lee are therefore both welcome. The former author is an American psychiatrist, the latter an Anglican priest. It is unexpected, therefore, to find the psychiatrist re-writing Freud to fit Catholic doctrine; and to find the priest adopting a rigid Freudian position, Oedipus complex and all!

Much of Dr Biddle's book is devoted to expounding his 'new system of depth psychology'; and the chief interest in this book lies in the description of a system that can be happily integrated with Roman Catholicism. Dr Lee's system, on the other hand, is plain, unvarnished Freud, and the chief interest of his book lies in discovering how such a mechanistic system can be shown to be compatible with Christian doctrine and practice.

But as Owen has pointed out, 'to expose the psychological or sociological origin of a belief is not at all the same thing as proving it false'; and Dr Lee has no difficulty in accepting that the sacrament of Holy Communion has the same psychological basis as smoking, kissing and cannibalism. He quotes Freud's fanciful account of the horde father being killed and eaten by his sons, as an 'excellent parable' to describe the unconscious psychological forces underlying Holy Communion.

Although these two books are interesting, and well worth reading, it seems to the reviewer that an empirical psychological study of the nature of religious experience would be more rewarding than the attempted reconciliation of rival dogmas.

RALPH HETHERINGTON

Education and Psychology (Mental Testing Number, 1954, nos. 4–5). Calcutta: Elm Press, 1954. Pp. 180. 3*s*.

It is difficult to assess the purpose underlying the publication of this special number. Side by side with accounts of the standardization of group tests in various Indian languages are several vehement and unoriginal denunciations of mental testing, including a particularly irresponsible example emanating from an English university ('...intelligence tests have been widely used to "prove" the mental superiority of the white over the coloured peoples...'). Also included are personal contributions by each of the joint editors, of which one, a review of a book on the same general theme, contains this remarkable *volte-face*: 'We are in agreement with [the author's] fundamental objections to the theory and practice of intelligence testing. But we see no reason why [he] objects to its use if they help us in proper educational guidance.'

The scientific papers are uneven in quality, but fairly wide in scope (e.g. R. G. Chatterjee presents 'An Evaluation of Reaction Time, Time Sense and Cutaneous Sensitivity in Three Pick-pockets'), and the writers show an earnest desire to make use of and apply the methods and findings of Western research.

BORIS SEMEONOFF

OTHER PUBLICATIONS RECEIVED

Max Gluckman. *The Judicial Process among the Barotse of Northern Nigeria.* Manchester: University Press. 1955. Pp. xxiii + 386. 37*s.* 6*d.* net.

A. Graham Ikin. *Live and Learn.* London: Epworth Press. 1955. Pp. 66. 6*s.* net.

S. J. Van Pelt. *Hypnotic Suggestion: its Role in Psychoneurotic and Psychosomatic Disorders.* Bristol: John Wright. 1955. Pp. 95. 8*s.* 6*d.*

J. E. Wallace Wallin. *The Odyssey of a Psychologist.* Published by the author, Wilmington, Delaware. 1955. Pp. xvii + 243. $3.00. (Cash with order, $2.50.)

H. R. Bowen. *The Business Enterprise as a subject for Research.* New York: Social Science Research Council (Pamphlet 11). 1955. Pp. viii + 103. $1.25.

W. M. Gillis. *The Poles in Wellington, New Zealand: a Phenomenological Study.* Wellington, N.Z.: Victoria University College Publications in Psychology, no. 5. 1954. Pp. 77. No price given.

M. Johansen. *An Introductory Study of Voluminal Form Perception.* Copenhagen: Nordisk Psykologi's Monograph Series. 1954. Pp. 51. No price given.

Helen M. Robinson. *Promoting Maximal Reading Growth among Able Learners.* Chicago: University Press; Cambridge University Press. Supplementary Educational Monographs, no. 81. 1954. Pp. viii + 191. 26*s.* net.

Alice S. Walker. *Pupil's School Records.* (Publications of the National Foundation for Educational Research in England and Wales, No. 8.) London: Newnes. 1955. Pp. xviii + 199. 21*s.*

The Ethics of Aristotle. Trans. by J. A. K. Thomson. Harmondsworth: Penguin Books. 1955. 2*s.* 6*d.*

The Young Worker: Education for Human Relations. London: Heinemann. 1955. Pp. 64. 3*s.* 6*d.*

International Journal on Alcohol and Alcoholism. Edited by E. M. Jellinek and H. Pullar-Strecker. Vol. I, no. 1. Oxford: Blackwell. 1955. 12*s.* 6*d.* net (Annual subscription: Post free 30*s.*).

The International Register of Current Team Research in the Social Sciences (1950–1952). Paris: UNESCO; London, H.M.S.O. 1955. Pp. 312. 22*s.* 6*d.*

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY
21 JUL 1962